

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2000, 17:34:18 ; Search time 17970.8 Seconds  
(without alignments)  
-1569.817 Million cell updates/sec

Title: US-08-852-495c-2\_COPY\_1\_29000  
Perfect score: 29000  
Sequence: 1 CACACACACACACACACA.....AACCTGCTCTCTGGGTTTC 29000

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 882769 seqs, -486395729 residues  
Total number of hits satisfying chosen parameters: 370290

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_ba3.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pl1.\*
- 8: gb\_pl2.\*
- 9: gb\_pr1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_vl.\*
- 17: em\_fun.\*
- 18: em\_hum1.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_om.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pl.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_vl.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_ba1.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*
- 44: gb\_htg6.\*

- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*
- 50: gb\_pl3.\*
- 51: gb\_pr5.\*
- 52: gb\_htg8.\*
- 53: gb\_htg9.\*
- 54: gb\_htg10.\*
- 55: gb\_htg11.\*
- 56: gb\_htg12.\*
- 57: gb\_htg13.\*
- 58: gb\_htg14.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query	Length	DB	ID	Description
C 1	87.2	0.3	103	9	HUMALCE221	M87896 Human carc1
C 2	87	0.3	107	9	HUMALCE162	M87924 Human carc1
C 3	87.2	0.3	108	10	HSLDLRN2	X05230 Human LDL-r
C 4	83.6	0.3	108	10	HSLDLRN2	X05230 Human LDL-r
C 5	79.8	0.3	108	10	HSLDLRD1	X05249 Human LDL-r
C 6	79.8	0.3	108	10	HSLDLRD2	X05251 Human LDL-r
C 7	80	0.3	108	11	HSU67803	U67803 Human small
C 8	75	0.3	103	9	HUMALCE221	M87896 Human carc1
C 9	75.2	0.3	108	11	HSU67804	U67804 Human small
C 10	74.2	0.3	108	9	HUMDID03M5	D16965 Human HepG2
C 11	73.6	0.3	108	10	HSLDL112	X05248 Human LDL-r
C 12	73.4	0.3	110	11	HSU67807	U67807 Human small
C 13	73	0.3	103	13	HS8IC8R	X57789 Human seque
C 14	73	0.3	108	10	HSLDLRD1	X05249 Human LDL-r
C 15	73	0.3	108	10	HSLDLRD2	X05251 Human LDL-r
C 16	72.4	0.2	101	10	S79560	S79560 HRX (intron
C 17	71.6	0.2	94	9	HUMHGAL	M13479 Human alpha
C 18	70.8	0.2	90	9	HUMDLRFL	K03555 Human low d
C 19	70.8	0.2	91	13	HUMUT8164A	L30244 Human STS U
C 20	69.8	0.2	108	13	G32614	G32614 A009K21 Hum
C 21	69.8	0.2	110	9	HUMALCE43	M87900 Human carc1
C 22	68.8	0.2	106	13	G32743	G32743 A009P31 Hum
C 23	69	0.2	108	11	HSU67803	U67803 Human small
C 24	68.8	0.2	108	11	HSU67808	U67808 Human small
C 25	68.4	0.2	95	13	HUMUT8002B	L30176 Human STS U
C 26	67.8	0.2	100	9	HUMGALNSA	D45223 Human GALNS
C 27	68	0.2	103	13	HS8IC8R	X57789 Human seque
C 28	68	0.2	104	9	HUMALCE272	M87899 Human carc1
C 29	68	0.2	107	9	HUMALCE162	M87924 Human carc1
C 30	67.8	0.2	108	13	G43535	G43535 WIAF-2393-S
C 31	67.2	0.2	85	10	HUMHIS1PR	M26162 Homo sapien
C 32	67.4	0.2	97	9	HUMDLRA2	M14180 Human low d
C 33	66.4	0.2	97	9	HUMDLRA1	M14178 Human low d
C 34	66.4	0.2	97	9	HUMDLRA2	M14180 Human low d
C 35	66	0.2	97	9	HUMDLRDJ	M14179 Human fam1
C 36	65.8	0.2	99	13	HUMUT7692A	L30306 Human STS U
C 37	66	0.2	100	13	G43536	G43536 WIAF-2394-S
C 38	66	0.2	100	13	G43538	G43538 WIAF-2396-S
C 39	66	0.2	110	9	HUMALCE43	M87900 Human carc1
C 40	65.6	0.2	80	9	HUMBRKFAE	M36135 Human alpha
C 41	65.6	0.2	107	11	HSU67806	U67806 Human small
C 42	65.2	0.2	79	10	S73203	S73203 ALL-1 {tand
C 43	65.4	0.2	110	11	HSU67807	U67807 Human small
C 44	64.6	0.2	95	10	HSSTHPKIB	X66361 H.sapiens m
C 45	64.8	0.2	96	4	NVHIS2A	J00950 Newt histon

ALIGNMENTS

```
RESULT 1
HUMALCE221/c HUMALCE221 103 bp ss-RNA PRI 15-APR-1994
LOCUS Human carcinoma cell-derived Alu RNA transcript, clone CE221.
DEFINITION M87896
ACCESSION M87896
VERSION M87896.1 GI:174874
KEYWORDS Alu repeat.
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 103)
AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
post-transcriptional selection of master sequences
JOURNAL J. Mol. Biol. (1992) In press
FEATURES
    source
        1..103
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /cell_line="Ntera2D1"
            /dev_stage="embryo"
            /sex="male"
            /tissue_type="carcinoma"
BASE COUNT 25 a 27 c 33 g 18 t
ORIGIN

Query Match 0.3%; Score 87.2; DB 9; Length 103;
Best Local Similarity 92.0%; Pred. No. 1.1e-05;
Matches 92; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 12183 CTGAGTGCATGGCGGAGCTTGGCTCAGACGACCTCGCCCTCCCGGTTCAAGCCAT 12242
Dy 103 CTGGAGTGCATGGCGGAGCTTGGCTCAGACGACCTCGCCCTCCCGGTTCAAGCGAT 44
Qy 12243 TCTCTGCTTACCTGCTCGGAGTACCTGGGATTACAGGCA 12282
Dy 43 TCTCTGCTTACCTGCTCGGAGTACCTGGGATTACAGGCA 4

RESULT 2
HUMALCE162/c HUMALCE162 107 bp ss-RNA PRI 15-APR-1994
LOCUS Human carcinoma cell-derived Alu RNA transcript, clone CE162.
DEFINITION M87924
ACCESSION M87924
VERSION M87924.1 GI:174871
KEYWORDS Alu repeat.
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 107)
AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
post-transcriptional selection of master sequences
JOURNAL J. Mol. Biol. (1992) In press
FEATURES
    source
        1..107
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /cell_line="Ntera2D1"
            /dev_stage="embryo"
            /sex="male"
            /tissue_type="carcinoma"
BASE COUNT 28 a 30 c 35 g 14 t
ORIGIN

Query Match 0.3%; Score 87; DB 9; Length 107;
Best Local Similarity 90.3%; Pred. No. 1.2e-05;
Matches 93; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 12183 CTGAGTGCATGGCGGAGCTTGGCTCAGACGACCTCGCCCTCCCGGTTCAAGCCAT 12242
Dy 103 CTGGAGTGCATGGCGGAGCTTGGCTCAGACGACCTCGCCCTCCCGGTTCAAGCGAT 44
Qy 12243 TCTCTGCTTACCTGCTCGGAGTACCTGGGATTACAGGCA 12282
Dy 43 TCTCTGCTTACCTGCTCGGAGTACCTGGGATTACAGGCA 4

RESULT 3
HSLDLRN2/c HSLDLRN2 108 bp DNA PRI 20-MAY-1992
LOCUS Human LDL-receptor gene intron 14 fragment (normal gene).
DEFINITION X05250
ACCESSION X05250
VERSION X05250.1 GI:34337
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Hovinga,J.R.,
Williamson,R. and Humphries,S.
TITLE The low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT See X05252 for deletion junction
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES
    source
        1..108
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /note="intron XIV fragment"
BASE COUNT 28 a 23 c 39 g 18 t
ORIGIN

Query Match 0.3%; Score 87.2; DB 10; Length 108;
Best Local Similarity 88.0%; Pred. No. 1.1e-05;
Matches 95; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 3618 CTGGGCTCACTGCAAGCTCTGCTCCCGGCTTATGCAATCTCATCTCAGCCTCAG 3677
Dy 108 CTGGGCTCACTGCAAGCTCTGCTCCCGGCTTATGCAATCTCATCTCAGCCTCAG 49
Qy 3678 AGTAGCTGGGACTACAGCGCGCCACACAGCCTGGCTAATTTT 3725
Dy 48 AGTAGCTGGGACTACAGCGCGCCACACAGCCTGGCTAATTTTGT 1

RESULT 4
HSLDLRN2 HSLDLRN2 108 bp DNA PRI 20-MAY-1992
LOCUS Human LDL-receptor gene intron 14 fragment (normal gene).
DEFINITION X05250
ACCESSION X05250
VERSION X05250.1 GI:34337
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Hovinga,J.R.,
Williamson,R. and Humphries,S.
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
```



COMMENT See X05252 for deletion junction  
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES  
Source 1..108  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..108  
/note="intron XIV fragment"  
28 a 23 c 39 g 18 t

BASE COUNT 28 a 23 c 39 g 18 t  
ORIGIN

Query Match 0.3%; Score 83.6; DB 10; Length 108;  
Best Local Similarity 86.8%; Pred. No. 4.5e-05;  
Matches 92; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 24483 AAAAATTAGCCAGGATGGTGGTGGCGGCTATATCCAGCTAATTTGGGAGGCTGAGGC 24542  
|||||  
Db 3 AAAAATTAGCCAGGATGGTGGTGGCGGCTATATCCAGCTAATTTGGGAGGCTGAGGC 62  
|||||

QY 24543 AGGAGAATTCCTGAACCTGGGAGGTGGAGGTTCGACCTGAGGCAAG 24588  
|||||  
Db 63 AGGAGAATTCCTGAACCTGGGAGGTGGAGGTTCGACCTGAGGCGAG 108  
|||||

RESULT 5  
LOCUS HSLDLRD1 108 bp DNA PRI 20-MAY-1992  
DEFINITION Human LDL-receptor mutated gene with intron 12 deletion junction.  
ACCESSION X05249  
VERSION X05249.1 GI:34335  
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 108)  
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,  
Williamson,R. and Humphries,S.  
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in  
the low-density-lipoprotein-receptor gene. A possible mechanism for  
the defect in a patient with familial hypercholesterolaemia  
Eur. J. Biochem. 164 (1), 77-81 (1987)

JOURNAL MEDLINE 87161901  
COMMENT \*source: hypercholesterol aemia  
See X05250 for corresponding normal gene sequence  
in the defective LDL-receptor gene the deletion occurred between two  
alu-repetitive sequences, that are in the same direction, the  
deletion eliminates exons 13 and 14 and changes the reading frame  
of the resulting spliced mRNA.  
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES  
Source Location/Qualifiers  
1..108  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="blood leukocytes from a patient with familial"  
1..108  
/note="intron XIV fragment"  
28 a 20 c 40 g 20 t

BASE COUNT 28 a 20 c 40 g 20 t  
ORIGIN

Query Match 0.3%; Score 79.8; DB 10; Length 108;  
Best Local Similarity 84.1%; Pred. No. 0.00021;  
Matches 90; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 3619 TTGGCTCACTGCAAGCTCTGCCGCTTATGCTCATCTCTCAGCTCCAGA 3678  
|||||  
Db 107 TGCCTCACACACCTCTGCTCTCTGCTTCAACCATTTCTCGCTCAGCTCCCA 48  
|||||

QY 3679 GTAGCTGGGACTACAGGCGCCGCCACCGCTGGCTAATTTTTT 3725  
|||||  
Db 47 GTAGCTGGGATTACAGGCACCTGCCACCACGCTGGCTAATTTTGT 1  
|||||

RESULT 7  
LOCUS HSU67803/3 108 bp RNA PRI 01-AUG-1997  
DEFINITION Human small cytoplasmic Alu transcript.  
ACCESSION U67803  
VERSION U67803.1 GI:2289917  
KEYWORDS Alu.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 108)  
AUTHORS Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.  
TITLE cDNAs derived from primary and small cytoplasmic Alu (scAlu)  
transcripts  
J. Mol. Biol. 271 (2), 222-234 (1997)  
JOURNAL MEDLINE 97415756  
COMMENT 2 (bases 1 to 108)  
REFERENCE 1 (bases 1 to 108)  
AUTHORS Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.  
TITLE Direct Submission  
JOURNAL Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The

Children's Hospital of Philadelphia, 1004F Abramson Research Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA

```

FEATURES
  source
    1..108
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="TscAlu2"
    1..108
      /note="scAlu"
      /rpt_family="Alu"
      /rpt_type-dispersed
  repeat_region
    23 a 39 c 30 g 16 t

BASE COUNT
ORIGIN
  23 a 39 c 30 g 16 t

Query Match
  0.3%; Score 80; DB 11; Length 108;
Best Local Similarity 89.6%; Pred. No. 0.0019;
Matches 86; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 3735 GTAGAGATGGGTTTCACCGTGTAGCCAGACGCTCTCGATCTCTTGACCTCTGTGATCC 3794
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 97 GTAGAGAGGGGTTTCACCTGTGTAGCCAGGATGCTCTCGATCTCTTGACCTCTGTGATCC 38
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 3795 GCCTGCTTGGCTCCCAAGTGTGGGATTACAGG 3830
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37 GCCCGCTCGGCTCCCAAGTGTGGGATTACAGG 2
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
HUMALCE221 HUMALCE221 103 bp ss-RNA PRI 15-APR-1994
LOCUS Human carcinoma cell-derived Alu RNA transcript, clone CE221.
DEFINITION M87896
ACCESSION M87896
VERSION M87896.1 GI:174874
KEYWORDS Alu repeat.
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 103)
AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
  post-transcriptional selection of master sequences
JOURNAL J. Mol. Biol. (1992) In press
FEATURES
  Location/Qualifiers
    1..103
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /cell_line="NTera2D1"
      /dev_stage="embryo"
      /sex="male"
      /tissue_type="carcinoma"
  BASE COUNT 25 a 27 c 33 g 18 t
ORIGIN
  25 a 27 c 33 g 18 t

Query Match
  0.3%; Score 75; DB 9; Length 103;
Best Local Similarity 84.8%; Pred. No. 0.0014;
Matches 84; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 24510 GCCTATATCCAGCTAATTTGGAGGTGAGCGAGGAGAAATGCTGNACTGGGAGGTG 24569
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5 GCCTATATCCAGCTACACGGAAGCTAAGCGAGGAGAAATGCTGNAACCGGAGCG 64
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 24570 GAGGTTCAGTCAGCCAGATACACACCATTCACCTCCAG 24608
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 GAGGTTCAGTCAGCCAGATACACACCATTCACCTCCAG 103
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
HSU67804/c HSU67804 108 bp RNA PRI 01-AUG-1997
LOCUS Human small cytoplasmic Alu transcript.
DEFINITION

```

```

ACCESSION U67804
VERSION U67804.1 GI:2289918
KEYWORDS Alu.
SOURCE human.
ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE cDNAs derived from primary and small cytoplasmic Alu (scAlu)
  transcripts
JOURNAL J. Mol. Biol. 271 (2), 222-234 (1997)
MEDLINE 97415756
REFERENCE 2 (bases 1 to 108)
AUTHORS Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The
  Children's Hospital of Philadelphia, 1004F Abramson Research
  Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA
  Location/Qualifiers
    1..108
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="TscAlu3"
    1..108
      /note="scAlu"
      /rpt_family="Alu"
      /rpt_type-dispersed
  repeat_region
    26 a 38 c 26 g 18 t

BASE COUNT
ORIGIN
  26 a 38 c 26 g 18 t

Query Match
  0.3%; Score 75.2; DB 11; Length 108;
Best Local Similarity 86.5%; Pred. No. 0.0013;
Matches 83; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 3735 GTAGAGATGGGTTTCACCGTGTAGCCAGACGCTCTCGATCTCTTGACCTCTGTGATCC 3794
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 97 GGAAGAGCGGGTTTCACCATGTTAGCCAGGATGCTCTCGATCTCTTGACCTCTGTGATCC 38
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 3795 GCCTGCTTGGCTCCCAAGTGTGGGATTACAGG 3830
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37 TCCCGCTTGGCTTCCCAAGTGTGGGATTACAGG 2
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
HUMD1D03M5/c HUMD1D03M5 108 bp mRNA PRI 04-FEB-1999
LOCUS Human HepG2 partial cDNA, clone hmdid03m5.
DEFINITION DI6965
ACCESSION DI6965
VERSION DI6965.1 GI:598552
KEYWORDS gene signature.
SOURCE Homo sapiens Male cell_line:HepG2 cDNA to mRNA, clone_11b:Kiseru.
ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
  Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Matoba,R.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-1993) to the DDBJ/EMBL/GenBank databases. Ryo
  Matoba, Osaka University, Institute for Molecular and Cellular Bio;
  1-3, Yamada-oka, Suita, Osaka 565, Japan
  (E-mail:matoba@inherit.imcb.osaka-u.ac.jp,
  Tel:81-6-877-5111(ex.3314), Fax:81-6-877-1922)
  2 (bases 1 to 108)
REFERENCE Matoba,R., Okubo,K., Hori,N., Fukushima,A. and Matsubara,K.
  The addition of 5'-coding information to a 3'-directed cDNA library
  improves analysis of gene expression
AUTHORS Gene 146 (2), 199-207 (1994)
TITLE Improves analysis of gene expression
JOURNAL 94357437
MEDLINE
COMMENT Submitted (21-Jul-1993) to DDBJ by:
  Ryo Matoba
  Molecular Microbiology and Genetics Lab.

```

Research Institute of Innovative Technology for the Earth 9-2  
 Kizugawadal Kizu-cho,  
 Soraku-gun, Kyoto  
 Japan, 619-02  
 Phone: 07747-5-2308  
 Fax: 07747-5-2321.

# FEATURES

source  
 1..108  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_line="HepG2"  
 /clone\_lib="Kiseru"  
 /sex="Male"

BASE COUNT 28 a 23 c 38 g 17 t 2 others  
 ORIGIN

Query Match 0.3%; Score 74.2; DB 9; Length 108;  
 Best Local Similarity 85.3%; Pred. No. 0.0019;  
 Matches 93; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 12200 GATCTGGCTACAGCAACCTCGGCTCCGGGTTCAAGCCATTCTCCTGCCTCAGCCCTC 12259  
 |||||  
 Db 108 GATCTGGCTACTGCAACCTCTGCTCCGGGNTCAAGCGACTCTCTGCCTCAGCCCTC 49

QY 12260 CGAGTAGCTGGGATTACAGGATCGCCGACGACACACCTGGGCTAAATTT 12308  
 |||||

Db 48 CTGAGTAGCTGGGATTACA-GCATGCGCCACACACNCTGGCTTTTAT 1

## RESULT 11

HSIDL112 108 bp DNA PRI 20-MAY-1992  
 LOCUS Human LDL-receptor gene intron 12 fragment (normal gene) LDL - low  
 DEFINITION density lipoprotein.

ACCESSION X05248  
 VERSION X05248.1 GI:34334

KEYWORDS Alu repetitive sequence; low density lipoprotein receptor;  
 repetitive sequence.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS Horsthemke, B., Beisiegel, U., Dunning, A., Hovinga, J.R.,  
 Williamson, R. and Humphries, S.  
 TITLE Unequal crossing-over between two alu-repetitive DNA sequences in  
 the low-density-lipoprotein-receptor gene. A possible mechanism for  
 the defect in a patient with familial hypercholesterolaemia

JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)  
 MEDLINE 87161901

COMMENT see X05249 for deletion junction  
 Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

# FEATURES

source  
 1..108  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 complement(<1..65)  
 /note="Alu repeat"

misc\_feature  
 intron  
 1..108  
 /note="intron XII fragment"  
 BASE COUNT 21 a 38 c 20 g 29 t  
 ORIGIN

Query Match 0.3%; Score 73.6; DB 10; Length 108;  
 Best Local Similarity 81.7%; Pred. No. 0.0024;  
 Matches 85; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 12204 TTGGCTCAGCAACCTCGGCTCCGGGTTCAAGCCATTCTCCTGCCTCAGCCCTCGGA 12263  
 |||||

Db 2 TGGCCTCACCACAACCTCGCTCTGGGTTCAAGCCATTCTCCTGCCTCAGCCCTCCTTA 61

QY 12264 GTAGCTGGGATTACAGGATCGCCGACGACACCTCGGCTAAATTT 12307  
 |||||

Db 62 GTAGCTGGGATTACAGGATGTGCCACCAGCCCGCGCTGATTTT 105

## RESULT 12

HSU67807/c 110 bp RNA PRI 01-AUG-1997  
 LOCUS Human small cytoplasmic Alu transcript.  
 DEFINITION U67807  
 ACCESSION U67807.1 GI:2289921  
 VERSION U67807.1  
 KEYWORDS Alu.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS Shaikh, T.H., Roy, A.M., Kim, J., Batzer, M.A. and Deininger, P.L.  
 TITLE CDNAs derived from primary and small cytoplasmic Alu (sAlu)  
 transcripts

JOURNAL J. Mol. Biol. 271 (2), 222-234 (1997)

## MEDLINE

97415756

## REFERENCE

2 (bases 1 to 110)

## AUTHORS

Shaikh, T.H., Kim, J., Batzer, M.A. and Deininger, P.L.

## TITLE

Direct Submission

## JOURNAL

Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The

Children's Hospital of Philadelphia, 1004F Abramson Research

Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA

## FEATURES

Location/Qualifiers

1..110

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="TscAlu6"

repeat\_region 1..110

/note="sAlu"

/rpt\_family="Alu"

/rpt\_type="dispersed"

BASE COUNT 26 a 39 c 24 g 21 t

ORIGIN

Query Match 0.3%; Score 73.4; DB 11; Length 110;

Best Local Similarity 83.8%; Pred. No. 0.0026;

Matches 83; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 2864 GTAGACATGGGTTTCACTATGTGGCCAGGCTAGTTTGAACCTCCTGACCTCAGTGAT 2923

|||

Db 99 GGAAGATGGGTTTCACTATGTGGCCAGGCTAGTTTGAACCTCCTGACCTCAGTGAT 40

|||

QY 2924 CCATTCTCTTGGCTCCCAAGTCTGGGATTACAGGC 2962

|||

Db 39 CCACCCACTTTGGCCCTCTCANAAGTCTGGGATTACAGGC 1

|||

## RESULT 13

HS81C8R 103 bp DNA STS 05-SEP-1991  
 LOCUS Human sequence tagged site 81C8R DNA from 19q13.  
 DEFINITION X57789

ACCESSION X57789.1 GI:23938

VERSION X57789.1

KEYWORDS STS; myotonic dystrophy.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 103)

AUTHORS Aldridge, F.L.

TITLE Direct Submission

JOURNAL Submitted (12-FEB-1991) F.L. Aldridge, ICI Pharmaceuticals,  
 Alderley Park, Macclesfield, Cheshire, SK10 4TG, UK

REFERENCE 2 (bases 1 to 103)

AUTHORS Butler, R., Riley, J.H., Ogilvie, D.J., Anand, R., Buxton, J.,  
 Davies, J., Johnson, K. and Markham, A.F.







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2000, 18:05:14 ; Search time 593.83 seconds  
(without alignments)  
12218.262 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_1\_29000  
Perfect score: 29000  
Sequence: 1 CACACACACACACACACA.....AACCTCGCTCTGGGTTTC 29000

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 433070

Minimum DB seq length: 10

Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68.4	0.2	108	1 X12095	Human biallelic po
2	66.6	0.2	100	1 X12085	Human biallelic po
3	66.6	0.2	100	1 X12086	Human biallelic po
4	65.4	0.2	86	1 V41231	Mouse embryonic ce
5	65	0.2	100	1 X12087	Human biallelic po
6	63.4	0.2	100	1 X12087	Human biallelic po
7	63.4	0.2	100	1 X12086	Human biallelic po
8	63.4	0.2	100	1 X12086	Human biallelic po
9	62.8	0.2	100	1 Q76490	Human genome fragm
10	62.2	0.2	108	1 X12095	Human biallelic po
11	60.8	0.2	100	1 T24892	Human gene signatu
12	58.6	0.2	103	1 T20927	Human gene signatu
13	57.8	0.2	103	1 T26213	Human gene signatu
14	57.8	0.2	108	1 T26828	Human gene signatu
15	56.4	0.2	84	1 T25848	Human gene signatu
16	56.2	0.2	108	1 T25009	Human gene signatu
17	54.8	0.2	87	1 T21566	Human gene signatu
18	54.4	0.2	110	1 T25260	Human gene signatu
19	53.6	0.2	69	1 Q29016	Probe to internal
20	53	0.2	106	1 Q95210	Simple tandem repe
21	52.6	0.2	110	1 T25260	Human gene signatu
22	52.2	0.2	65	1 T25588	Human gene signatu
23	52	0.2	93	1 T25688	Human gene signatu
24	52.2	0.2	103	1 T20927	Human gene signatu
25	52	0.2	108	1 T26828	Human gene signatu
26	51.4	0.2	70	1 N60231	Normal chromosome
27	51.2	0.2	100	1 T24892	Human gene signatu
28	51.4	0.2	102	1 T20743	Human gene signatu
29	51.6	0.2	110	1 T26288	Human gene signatu
30	50.6	0.2	99	1 T23728	Human gene signatu
31	49.8	0.2	91	1 T25854	Human gene signatu
32	50	0.2	99	1 T20931	Human gene signatu
33	49.6	0.2	93	1 T22572	Human gene signatu
34	49.2	0.2	108	1 T25009	Human gene signatu

c 35	48.4	0.2	93	1 T22572	Human gene signatu
c 36	48.6	0.2	97	1 T26728	Human gene signatu
c 37	48.2	0.2	93	1 T24259	Human gene signatu
c 38	47.6	0.2	69	1 T24175	Human gene signatu
c 39	47.8	0.2	92	1 T25052	Human gene signatu
c 40	47.4	0.2	65	1 T24893	Human gene signatu
c 41	47.6	0.2	95	1 Q75099	Plasmid pOKS18a c
c 42	47.6	0.2	109	1 T23895	Human gene signatu
c 43	47.2	0.2	69	1 Q29016	Probe to internal
c 44	47.2	0.2	85	1 T24033	Human gene signatu
c 45	47	0.2	94	1 T26403	Human gene signatu

ALIGNMENTS

RESULT 1	
X12095	
ID X12095 standard; DNA; 108 BP.	
AC X12095;	
DT 30-MAR-1999 (first entry)	
DE Human biallelic polymorphic DNA fragment TIGR-A003M18a.	
KW Polymorphism: biallelic; human; forensic: paternity testing; disease;	
KW detection; phenotypic typing; characteristic; infection; hereditary;	
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;	
KW treatment; marker; ss.	
OS Homo sapiens.	
PN W09820165-A2.	
PD 14-MAY-1998.	
PF 05-NOV-1997; U20313.	
PR 06-NOV-1996; US-030455.	
PA (WRED ) WHITEHEAD INST BIOMEDICAL RES.	
PI Hudson T, Lander ES, Wang D;	
DR WPI; 98-286974/25.	
PT New isolated nucleic acid segments from the human genome - used for	
PT determining polymorphic forms for use in e.g. forensics, paternity	
PT testing or phenotypic typing for disease	
PS Claim 1: Page 219; 310pp; English.	
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic	
CC markers which have been isolated using the primers represented in	
CC X09121-X10268. The base occupying the polymorphic site is indicated by	
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in	
CC methods for determining polymorphic forms in an individual for use in	
CC e.g. forensics, paternity testing or for phenotypic typing for diseases	
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,	
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial	
CC hypercholesterolemia, polycystic kidney disease, hereditary	
CC spherocytosis, von Willebrand's disease, tubercous sclerosis, hereditary	
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos	
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,	
CC autoimmune diseases, inflammation, cancer, diseases of the nervous	
CC system, infection by pathogenic microorganisms, and characteristics such	
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,	
CC endurance, fertility, and susceptibility or receptivity to particular	
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid	
CC segments can also be used to produce medicaments for the treatment or	
CC prophylaxis of such diseases.	
CC Sequence 108 BP; 19 A; 23 C; 28 G; 37 T;	
SQ	

Query Match	0.2%;	Score 68.4;	DB 1;	Length 108;
Best Local Similarity	82.4%;	Pred. No. 0.031;		
Matches 89;	Conservative 1;	Mismatches 17;	Indels 1;	Gaps 1;
QY 2854	TCGATTTTGTAGTAGATGGGTTTCACATATGTTGGCCAGGCTAGTTTGGAACTCTGAC	2913		
Db 1	TCGCTTTTGTAGAGATGAGGTTTCCTTGTGTCAGAGATGGTCTCGAACTCTCTGAC	60		
QY 2914	CTCCAGTGATCCATTCATTCATGGCCCTCCC-AAAGTCTGGGATTACAG	2960		
Db 61	TTCAAGTGATCCGTCGTGCCTTGGCCTCCCAAAAGTGCTGGGATTATAG	108		

```
RESULT 2
X12085/c
ID X12085 standard; DNA; 100 BP.
AC X12085;
DT
DE Human biallelic polymorphic DNA fragment EST98276c.
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW treatment; marker; ss.
OS Homo sapiens.
PN W09820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; U20313.
PR 06-NOV-1996; US-030455.
PA (WHEAT ) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
DR WPI; 98-286974/25.
PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
PS Claim 1; Page 218; 310pp; English.
CC X12069-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberculous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC segments can also be used to produce medicaments for the treatment or
CC prophylaxis of such diseases.
SQ Sequence 100 BP; 22 A; 25 C; 22 G; 30 T;

Query Match 0.2%; Score 66.6; DB 1; Length 100;
Best Local Similarity 78.8%; Pred. No. 0.058;
Matches 78; Conservative 1; Mismatches 20; Indels 0; Gaps 0;

Qy 22975 GTGGCTCATGCTGTAATCCAGACATCTTGAGAGGCTGAAGAGGAGGATCGTTGAGTC 23034
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99 GTGACTCACACCTATATCTCGGCACCTTAGAGGCTTAGGAGGAGGATGTTTGAAC 40

Qy 23035 CGGGAGTTCAAGACATCTCGGCAACACACGAGACCC 23073
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 39 CAGGAGCTCAAGACATCTCGGCAACACATAGCAAGATC 1

RESULT 4
X1231/c
ID V41231 standard; cDNA; 86 BP.
AC V41231;
DT 01-OCT-1998 (first entry)
DE Mouse embryonic cell EST 13-4 nucleotide sequence.
KW Embryonic stem cell; ESC; non-primate; mouse; EST; human;
KW developmental gene; transgenic animal; reporter gene; ss.
OS Mus sp.
PN W09823633-A1.
PD 04-JUN-1998.
PF 25-NOV-1997; U22335.
PR 27-NOV-1996; US-032510.
PA (CORR ) CORNELL RES FOUND INC.
PI Holtschu DL, Mark WH;
DR WPI; 98-322656/28.
PT Screening for human developmental genes - by trapping in murine
PT embryonic stem cells and analysing differential expression in vitro,
PT selecting homologous non-human primate gene and using it to isolate
PT human gene
PS Claim 37; Page 18; 60pp; English.
CC Sequences shown in V41230 to V41247 represent nucleotide sequences of
CC mouse EST from tagged cDNA clones. These are used in the method of the
CC invention of screening for human developmental genes. The method
CC comprises inserting a promoterless reporter gene into a non-primate
CC mammalian embryonic stem cell (ESC) genome and identifying cellular
CC transcripts that encode the reporter gene product. Fragments of genes
CC encoding these transcripts are cloned and sequenced. A gene encoding a
CC transcript that includes unknown sequences is selected and expression
CC level of the gene encoding the transcript, or part of it, in different
CC cell types and/or different developmental stages is detected. A gene
CC showing differential expression is selected and expression levels of a
```



CC homologous non-human primate gene, in different cell types and/or at  
 CC different developmental stages, using the non-primate transcript as  
 CC probe is detected. A homologous gene having the same pattern of  
 CC differential expression is selected and the non-primate gene, or part of  
 CC it is used to identify the homologous human gene. The ESC transcripts  
 CC identified by this method are used to generate transgenic animals  
 CC selected from rats, hamsters, rabbits, dogs, pigs, horses, cows, monkey,  
 CC baboon or chimpanzee for study of gene function. The method provides  
 CC rapid and large scale screening for human developmental genes, and  
 CC eliminates the need to analyse reporter gene expression in embryos.  
 SQ Sequence 86 BP; 16 A; 28 C; 30 G; 12 T;

Query Match 0.2%; Score 65.4; DB 1; Length 86;  
 Best Local Similarity 86.7%; Pred. No. 0.086; Mismatches 0; Gaps 0;  
 Matches 72; Conservative 0; Indels 11;

QY 9607 TGACTCTCGTCTTCTAGGACAGACGCGCTGGATTTAGGAGACGCGCCCTGA 9666

DB 84 TGGCTCTCGTCTTCTTGGCAGCAGCGCGCTGGATTTGGCAGACGCGCGCTGC 25

QY 9667 GCAATGGTCACCGCGCTAGCAG 9689

DB 24 GCGATGGTCACGCGCCAGCAG 2

RESULT 5  
 ID X12087/C  
 AC X12087 standard; DNA; 100 BP.  
 DT 30-MAR-1999 (first entry)  
 DE Human biallelic polymorphic DNA fragment EST98276a.  
 KW Polymorphism; biallelic; human; forensic; paternity testing; disease;  
 KW detection; phenotypic typing; characteristic; infection; hereditary;  
 KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;  
 KW treatment; marker; ss.  
 OS Homo sapiens.  
 PN W09820165-A2.  
 PD 14-MAY-1998.  
 PF 05-NOV-1997; U20313.  
 PR 06-NOV-1996; US-030455.  
 PA (WRED ) WHITEHEAD INST BIOMEDICAL RES.  
 PI Hudson T, Lander ES, Wang D;  
 DI WPI; 98-286974/25.  
 PT New isolated nucleic acid segments from the human genome - used for  
 PT determining polymorphic forms for use in e.g. forensics, paternity  
 PT testing or phenotypic typing for disease  
 PS Claim 1; Page 219; 310pp; English.

CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
 CC markers which have been isolated using the primers represented in  
 CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
 CC methods for determining polymorphic forms in an individual for use in  
 CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
 CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
 CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
 CC hypercholesterolemia, polycystic kidney disease, hereditary  
 CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary  
 CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
 CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
 CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
 CC system, infection by pathogenic microorganisms, and characteristics such  
 CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
 CC endurance, fertility, and susceptibility or receptivity to particular  
 CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
 CC segments can also be used to produce medicaments for the treatment or  
 SQ prophylaxis of such diseases.  
 Sequence 100 BP; 21 A; 25 C; 22 G; 31 T;

Query Match 0.2%; Score 65; DB 1; Length 100;  
 Best Local Similarity 77.8%; Pred. No. 0.099; Mismatches 21; Indels 0; Gaps 0;  
 Matches 77; Conservative 1;

QY 22975 GTGGCTCATGCTTAATCCAGCACTTTGAGAGGCTGAAGAGGAGGATCGCTTGATC 23034

DB 99 GTGACTCACACTATATCTGCGCACTTTAGGAGGCTTAGGAGGAGGATTTTGAAC 40

QY 23035 CGGGAGTTTCAAGCATCTCTGGGCAACACAGCGAGACCC 23073

DB 39 CAGGAGCTCAGACCAKCKCTGGGAAACATAGCAAGACTC 1

RESULT 6

ID X12087 standard; DNA; 100 BP.  
 AC X12087;  
 DT 30-MAR-1999 (first entry)

DE Human biallelic polymorphic DNA fragment EST98276a.  
 KW Polymorphism; biallelic; human; forensic; paternity testing; disease;  
 KW detection; phenotypic typing; characteristic; infection; hereditary;  
 KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;  
 KW treatment; marker; ss.

OS Homo sapiens.

PN W09820165-A2.

PD 14-MAY-1998.

PF 05-NOV-1997; U20313.

PR 06-NOV-1996; US-030455.

PA (WRED ) WHITEHEAD INST BIOMEDICAL RES.

PI Hudson T, Lander ES, Wang D;

DI WPI; 98-286974/25.

PT New isolated nucleic acid segments from the human genome - used for  
 PT determining polymorphic forms for use in e.g. forensics, paternity  
 PT testing or phenotypic typing for disease

PS Claim 1; Page 219; 310pp; English.

CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
 CC markers which have been isolated using the primers represented in  
 CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
 CC methods for determining polymorphic forms in an individual for use in  
 CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
 CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
 CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
 CC hypercholesterolemia, polycystic kidney disease, hereditary  
 CC spherocytosis, von Willebrand's disease, tuberous sclerosis, Ehlers-Danlos  
 CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
 CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
 CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
 CC system, infection by pathogenic microorganisms, and characteristics such  
 CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
 CC endurance, fertility, and susceptibility or receptivity to particular  
 CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
 CC segments can also be used to produce medicaments for the treatment or  
 SQ prophylaxis of such diseases.  
 Sequence 100 BP; 21 A; 25 C; 22 G; 31 T;

Query Match 0.2%; Score 63.4; DB 1; Length 100;  
 Best Local Similarity 76.8%; Pred. No. 0.17; Mismatches 22; Indels 0; Gaps 0;  
 Matches 76; Conservative 1;

QY 17671 GGTCTTACTATGTTCCCGAGGCTGCTCAAACTCCTGGCTTAAGTATCTCTGCC 17730

DB 1 GAGCTCTTGCTATGTTTCCCGAGGTTGCTTGTAGCTCTGTTTCAACAACATCTCTCTCC 60

QY 17731 TCAGCTCCCAATTTCTGGGATTACTAGTGAGTCAC 17769

DB 61 TAAGCTCTCTAAAGTCCCGAGGATTATAGGTGTGAGTCAC 99

RESULT 7

ID X12085 standard; DNA; 100 BP.  
 AC X12085;  
 DT 30-MAR-1999 (first entry)

DE Human biallelic polymorphic DNA fragment EST98276c.

KW	Polymorphism; biallelic; human; forensic; paternity testing; disease; detection; phenotypic typing; characteristic; infection; hereditary; autoimmune disease; cancer; inflammation; drug; therapy; medicament; treatment; marker; ss.
KW	Homo sapiens.
OS	WO9820165-A2.
PN	14-MAY-1998.
PD	05-NOV-1997; U20313.
PF	05-NOV-1996; US-030455.
PR	(WHED ) WHITEHEAD INST BIOMEDICAL RES.
PA	Hudson T, Lander ES, Wang D;
PI	WPI; 98-286974/25.
DR	New isolated nucleic acid segments from the human genome - used for determining polymorphic forms for use in e.g. forensics, paternity testing or phenotypic typing for disease
PT	Claim 1: Page 218; 310pp; English.
PS	X10269-X12937 are human DNA fragments which contain biallelic polymorphic markers which have been isolated using the primers represented in X09121-X10268. The base occupying the polymorphic site is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments can be used in methods for determining polymorphic forms in an individual for use in e.g. forensics, paternity testing or for phenotypic typing for diseases
CC	such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial hypercholesterolemia, polycystic kidney disease, hereditary haemochromatosis, von Willebrand's disease, tuberous sclerosis, hereditary spherocytosis, telangiectasia, familial colonic polyposis, Ehlers-Danlos syndrome, osteogenesis imperfecta, acute intermittent porphyria, autoimmune diseases, inflammation, cancer, diseases of the nervous system, infection by pathogenic microorganisms, and characteristics such as longevity, appearance (e.g. baldness, obesity), strength, speed, endurance, fertility, and susceptibility or receptivity to particular drugs or therapeutic treatments. The isolated polymorphic nucleic acid segments can also be used to produce medicaments for the treatment or prophylaxis of such diseases.
CC	Sequence 100 BP; 22 A; 25 C; 22 G; 30 T;
SQ	
Query Match 0.2%; Score 63.4; DB 1; Length 100;	
Best Local Similarity 76.8%; Pred. No. 0.17; 22; Indels 0; Gaps 0	
Matches 76; Conservative 1; Mismatches	
QY	17671 GGCTCTACTATGTTGCCAGCGTGTCTCAAACTCTGGCGTTAAAGTGATCCTCTGCC 17730
Db	TTTTCTGTATGTTTCCAGGATGGTCTGAGCTCTGGTTTCAACATCCTCTTCC 60
QY	17731 TCAGCCTCCCAAAATGTTGGGATTTACTAGTGTGAGTCAC 17769
Db	TTTTCTGTATGTTTCCAGGATGGTCTGAGCTCTGGTTTCAACATCCTCTTCC 60
RESULT 8	
X12086	12086 standard; DNA; 100 BP.
ID	X12086;
AC	30-MAR-1999 (first entry)
DE	Human biallelic polymorphic DNA fragment EST98276b.
KW	Polymorphism; biallelic; human; forensic; paternity testing; disease; detection; phenotypic typing; characteristic; infection; hereditary; autoimmune disease; cancer; inflammation; drug; therapy; medicament; treatment; marker; ss.
KW	Homo sapiens.
OS	WO9820165-A2.
PN	14-MAY-1998.
PD	05-NOV-1997; U20313.
PF	05-NOV-1996; US-030455.
PR	(WHED ) WHITEHEAD INST BIOMEDICAL RES.
PA	Hudson T, Lander ES, Wang D;
PI	WPI; 98-286974/25.
DR	New isolated nucleic acid segments from the human genome - used for determining polymorphic forms for use in e.g. forensics, paternity testing or phenotypic typing for disease
PT	Claim 1: Page 219; 310pp; English.

```
QY 9645 GTTAGGAGGACGC 9658
Db 14 GTTGGCAGGACGC 1

RESULT 10
X12095/c
ID X12095 standard; DNA; 108 BP.
AC X12095;
DT 30-MAR-1999 (first entry)
DE Human biallelic polymorphic DNA fragment TIGR-A003M18a.
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW treatment; marker; ss.
OS Homo sapiens.
PN W09820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; U20313.
PR 06-NOV-1996; US-030455.
PI (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
WPI; 98-286974/25..
DR New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
PS Claim 1: Page 219; 310pp; English.
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC X09121-X10268. The base occupying the polymorphic site is indicated by
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC spherocytosis, von Willebrand's disease, tuberculous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria.
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC segments can also be used to produce medicaments for the treatment or
CC prophylaxis of such diseases.
CC Sequence 108 BP; 19 A; 23 C; 28 G; 37 T;

Query Match 0.2%; Score 62.2; DB 1; Length 108;
Best Local Similarity 79.4%; Pred No. 0.26;
Matches 85; Conservative 1; Mismatches 19; Indels 2; Gaps 1;

QY 24378 TGTATTCCAGCATT--GGAGGCAGAGCGCGGAGATCACTTGAGGTGGGAGTTGCA 24435
Db 107 TATAATCCAGCATTGTTGGAGGCGGCGGAGCGGATCACTTGAAGTCAGGAGTTCCA 48
QY 24436 GACTAGCCTGGCCACATGATGAAACCCCTCTCTACTAAAAATACA 24482
Db 47 GACCATCTGTGGCCACAYAGGAACCCCTCTCTCTACAAAAAAGACA 1

RESULT 11
T24892/c
ID T24892 standard; cDNA to mRNA; 100 BP.
AC T24892;
DT 05-NOV-1996 (first entry)
DE Human gene signature HUMGS06998.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.

Query Match 0.2%; Score 60.8; DB 1; Length 100;
Best Local Similarity 74.7%; Pred No. 0.41;
Matches 74; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 3559 TTTTFTTTTTCAGAGGAGTCTAGCTCTGTCGCCAGGCTGGAGTGGCACCATC 3618
Db 100 TTTGTTTGTGTTTCAACAGAGTGTCACTCTGTACCCAGGCGAGGTGCAATC 41
QY 3619 TTGGCTCACTGCAAGCTCTGCTCCCGGCTTTATGCCAT 3657
Db 40 TCAGCTNATTTGCAAAATCTGCTCCAGGTTCAAGCGAT 2

RESULT 12
T20927/c
ID T20927 standard; cDNA to mRNA; 103 BP.
AC T20927;
DT 24-JUL-1996 (first entry)
DE Human gene signature HUMGS02180.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PR 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PI (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
WPI; 95-206931/27.
DR Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1: Page 758-759; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in 191001-726837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
```

```
PD 01-JUN-1995.
PR 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PI (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
WPI; 95-206931/27.
DR Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1: Page 1720; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in 191001-726837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
CC Sequence 100 BP; 28 A; 22 C; 25 G; 22 T;

Query Match 0.2%; Score 60.8; DB 1; Length 100;
Best Local Similarity 74.7%; Pred No. 0.41;
Matches 74; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 3559 TTTTFTTTTTCAGAGGAGTCTAGCTCTGTCGCCAGGCTGGAGTGGCACCATC 3618
Db 100 TTTGTTTGTGTTTCAACAGAGTGTCACTCTGTACCCAGGCGAGGTGCAATC 41
QY 3619 TTGGCTCACTGCAAGCTCTGCTCCCGGCTTTATGCCAT 3657
Db 40 TCAGCTNATTTGCAAAATCTGCTCCAGGTTCAAGCGAT 2

RESULT 12
T20927/c
ID T20927 standard; cDNA to mRNA; 103 BP.
AC T20927;
DT 24-JUL-1996 (first entry)
DE Human gene signature HUMGS02180.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PR 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PI (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
WPI; 95-206931/27.
DR Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1: Page 758-759; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in 191001-726837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
```

```
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 103 BP; 22 A; 27 C; 21 G; 31 T;

Query Match 0.2%; Score 58.6; DB 1; Length 103;
Best Local Similarity 73.7%; Pred. No. 0.85;
Matches 73; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 1461 ATGGCGAACCCTCTCTACTAAATACAAAATTAGCTGGGCATGGTGTGCTTC 1520
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 99 ATGGGAATACTGTCCTCTAAATACAAAATTAGCTGGGCATGGTGTGACACAC 40
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1521 TGTGTCCAGCTACTCTGGGAGGCTGAGGCTGAAGAATC 1559
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 39 TGTAGCCACACTTCTGGGAGTGGAGTGGGAGGATC 1
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
T26213/c
ID T26213 standard; cDNA to mRNA; 103 BP.
AC T26213; 1996 (first entry)
DE Human gene signature HUMGS08452.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-Al.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues.
PS Claim 1; Page 2029; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 103 BP; 33 A; 21 C; 25 G; 23 T;

Query Match 0.2%; Score 57.8; DB 1; Length 103;
Best Local Similarity 73.3%; Pred. No. 1.2;
Matches 74; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 13746 TTTTCTTTTGTGACAGAGAAATTTGCTGTGCCCCAGGCTGGGTGCGAGTGCCAA 13805
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 102 TTTTCTTTTAAAGACATGTTCTTACTCTGTGCCCCAGGCTGGGTGCGAGTGCCAA 43
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Qy 13806 TCTCGGCTCACTGCAACCTCTGCCTCCAGGTTTCTAGCAAT 13846
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 42 TCATAGTCACTGTATACACCAAACTCTGGACTCAAGTGAT 2
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
T26828
ID T26828 standard; cDNA to mRNA; 108 BP.
AC T26828;
DE Human gene signature HUMGS09078.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-Al.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues.
PS Claim 1; Page 2182; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 108 BP; 18 A; 33 C; 23 G; 28 T;

Query Match 0.2%; Score 57.8; DB 1; Length 108;
Best Local Similarity 82.3%; Pred. No. 1.2;
Matches 65; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 3774 GATCTCTTGACCTTCTGATCCGCTGCTTCCCAAGTGTCCAGGTGT 3833
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1 GATCTCTTGACCTTCTGATCCGCTGCTTCCCAAGTGTCCAGGTGT 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 3834 GAGCCACCATGCCGGCTG 3852
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 61 GAGCCACCACGCCGGCTG 79
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
T25848/c
ID T25848 standard; cDNA to mRNA; 84 BP.
AC T25848;
DE Human gene signature HUMGS08078.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-Al.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
```

PA (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI: 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1: Page 1942; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 84 BP; 33 A; 17 C; 15 G; 19 T;

Query Match 0.2%; Score 56.4; DB 1; Length 84;  
Best Local Similarity 80.5%; Pred. No. 1.8;  
Matches 66; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 17641 TAATTTTAAAGGCTTTTGTAGAGATGGGCTCTACTATGTTGCCAGGCTGGTCTC 17700  
Db ||||||| ||| ||||||| ||| ||||| ||| ||||||| |||||  
QY 17701 AAATCCTGGGCTTAAGTGATC 17722  
Db ||||||| ||||| |||||  
QY 17701 AAATCCTGGGCTTAAGTGATC 17722  
Db ||||||| ||||| |||||  
QY 17701 AAATCCTGGGCTTAAGTGATC 17722  
Db ||||||| ||||| |||||

Search completed: June 18, 2000, 01:53:51  
Job time: 305742 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2000, 13:46:22 ; Search time 8513.5 Seconds  
(without alignments)  
13806.716 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_1\_29000  
Perfect score: 29000  
Sequence: 1 CACACACACACACACACACA.....AACCTCTGCCTCCTGGGTTC 29000

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 156056

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:

- 1: em\_est1:\*
- 2: em\_est2:\*
- 3: em\_est3:\*
- 4: em\_est4:\*
- 5: em\_est5:\*
- 6: em\_est6:\*
- 7: em\_est7:\*
- 8: em\_est8:\*
- 9: em\_est9:\*
- 10: em\_est10:\*
- 11: em\_est11:\*
- 12: em\_est12:\*
- 13: em\_est13:\*
- 14: em\_est14:\*
- 15: em\_est15:\*
- 16: em\_est16:\*
- 17: em\_est17:\*
- 18: em\_est18:\*
- 19: em\_est19:\*
- 20: gb\_est1:\*
- 21: gb\_est2:\*
- 22: gb\_est3:\*
- 23: gb\_est4:\*
- 24: gb\_est5:\*
- 25: gb\_est6:\*
- 26: gb\_est7:\*
- 27: gb\_est8:\*
- 28: gb\_est9:\*
- 29: gb\_est10:\*
- 30: gb\_est11:\*
- 31: gb\_est12:\*
- 32: gb\_est13:\*
- 33: gb\_est14:\*
- 34: gb\_est15:\*
- 35: gb\_est16:\*
- 36: gb\_est17:\*
- 37: gb\_est18:\*
- 38: gb\_est19:\*
- 39: gb\_est20:\*
- 40: gb\_est21:\*
- 41: gb\_est22:\*
- 42: gb\_est23:\*
- 43: gb\_est24:\*
- 44: gb\_est25:\*

- 45: gb\_est26:\*
- 46: gb\_est27:\*
- 47: gb\_est28:\*
- 48: gb\_est29:\*
- 49: gb\_est30:\*
- 50: gb\_est31:\*
- 51: gb\_est32:\*
- 52: em\_est20:\*
- 53: em\_est21:\*
- 54: em\_est22:\*
- 55: em\_est23:\*
- 56: em\_est24:\*
- 57: em\_est25:\*
- 58: em\_est26:\*
- 59: gb\_est33:\*
- 60: gb\_est34:\*
- 61: gb\_est35:\*
- 62: gb\_est36:\*
- 63: gb\_est37:\*
- 64: gb\_est38:\*
- 65: em\_est27:\*
- 66: em\_est28:\*
- 67: em\_est29:\*
- 68: em\_est30:\*
- 69: gb\_est39:\*
- 70: gb\_est40:\*
- 71: gb\_est41:\*
- 72: gb\_est42:\*
- 73: gb\_est43:\*
- 74: gb\_est44:\*
- 75: em\_est31:\*
- 76: em\_est32:\*
- 77: em\_est33:\*
- 78: em\_est34:\*
- 79: gb\_est45:\*
- 80: gb\_est46:\*
- 81: gb\_est47:\*
- 82: gb\_gss1:\*
- 83: gb\_gss2:\*
- 84: gb\_gss3:\*
- 85: gb\_gss4:\*
- 86: em\_gss1:\*
- 87: em\_gss2:\*
- 88: em\_gss3:\*
- 89: em\_gss4:\*
- 90: gb\_gss5:\*
- 91: gb\_gss6:\*
- 92: gb\_gss7:\*
- 93: gb\_gss8:\*
- 94: gb\_gss9:\*
- 95: em\_gss5:\*
- 96: em\_gss6:\*
- 97: em\_gss7:\*
- 98: em\_gss8:\*
- 99: em\_gss9:\*
- 100: em\_gss10:\*
- 101: em\_gss11:\*
- 102: gb\_gss10:\*
- 103: gb\_gss11:\*
- 104: em\_gss12:\*
- 105: gb\_gss12:\*
- 106: gb\_gss13:\*
- 107: gb\_gss14:\*
- 108: gb\_gss15:\*
- 109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query





clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 727 Std Error: 0.00

Seq primer: M13RP1

High quality sequence stop: 68.

Location/Qualifiers

#### FEATURES

source

1. .95  
/organism="Homo sapiens"  
/db\_xref="GDB:479462"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1217301"  
/clone\_lib="Soares fetal liver spleen INFLS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACTGGGAAGTAATAAAGATCTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

#### BASE COUNT

31 a 23 c 26 g 12 t 3 others

#### Query Match

Best Local Similarity 0.3%; Score 92; DB 21; Length 95;

Matches 92; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 10297 TCTTTTAGATGCTGACCGACCAAGTCGTCTCCGCCCGCCGAGAGGGCTCCAGAG 10356  
Db 1 TCTTTTAGATGCTGACCGACCAAGTCGTCTCCGCCCGCCGAGAGGGCTCCAGAG 60

QY 10357 GCAGTGACCAAGCGCAGAGAAGATGCGAAGAA 10391

Db 61 GCAGTGACCAAGCGCAGAGAAGANNCAGAA 95

#### RESULT

3

LOCUS AI832832 105 bp mRNA EST 13-JUL-1999

DEFINITION at72g09.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone IMAGE:2377600 3' similar to contains Alu repetitive element; contains element MER22 repetitive element ;, mRNA sequence.

ACCESSION AI832832

VERSION AI832832.1 GI:5454812

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

1 (bases 1 to 105)

AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,

Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,

Thesling, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.

WashU-NCI human EST Project

Unpublished (1997)

JOURNAL On Dec 20, 1995 this sequence version replaced gi:1133644.

COMMENT Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -400P from Gibco.

#### FEATURES

source

1. .105

Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2377600"  
/clone\_lib="Barstead colon HPLRB7"  
/sex="male"  
/dev\_stage="adult, age 25"  
/lab\_host="DH10B (phage resistant)"  
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACGAATCTGAAGTGGAGCGCCGCTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [5', AATTCACCTAGTAAT 3' and 5', ATTACTAGTG 3'], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead."

BASE COUNT 17 a 35 c 27 g 26 t

#### ORIGIN

Query Match 0.3%; Score 92.2; DB 61; Length 105;

Best Local Similarity 92.4%; Pred. No. 0.061; Mismatches 8; Indels 0; Gaps 0;

Matches 97; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 12154 GAGACGAGCTTCTCTCTGTTTCCAGGCTGGAGTGCATGCGCGATCTTTGGCTCACA 12213

Db 1 GAGACGAGCTTCTCTCTGTTTCCAGGCTGGAGTGCATGCGCGATCTTTGGCTCACC 60

QY 12214 GCAACCTCCGCTCCCGGGTTCAGGCCATTCTCTCCCTCAGCCCT 12258

Db 61 GCAACCTCCGCTCCCGGGTTCAGGCCATTCTCTCCCTCAGCCCT 105

#### RESULT

4

LOCUS AA807640

DEFINITION AA807640

ACCESSION AA807640

VERSION AA807640.1 GI:2877108

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

1 (bases 1 to 103)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Unpublished (1997)

JOURNAL On Jan 19, 1998 this sequence version replaced gi:2151346.

COMMENT Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone distribution: NCI-CGAP clone distribution

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 774 Std Error: 0.00

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 87.

#### FEATURES

source

1. .103

Location/Qualifiers

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1255473"

/clone\_lib="NCI-CGAP\_GC3"

/tissue\_type="pooled germ cell tumors"

/lab\_host="DH10B"  
/note="vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from 3 pooled  
germ cell tumors, and was then primed with a Not I -  
oligo(dT) primer. Double-stranded cDNA was ligated to Eco  
RI adaptors (Pharmacia), digested with Not I and cloned  
into the Not I and Eco RI sites of the modified pT7T3  
vector. Library is not normalized. Library was  
constructed by Bento Soares and M. Fatima Bonaldo. "  
BASE COUNT 19 a 27 c 30 g 27 t  
ORIGIN

Query Match 0.3%; Score 87.6; DB 38; Length 103;  
Best Local Similarity 91.2%; Pred. No. 0.21;  
Matches 93; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
Qy 3734 AGTAGAGATGGGTTTACCGTGTAGCCAGAGCGTCTCGATCTCTGACCTTCTGATC 3793  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 2 AGTAGAGATGGGTTTACCGTGTAGCCAGAGCGTCTCGATCTCTGACCTTCTGATC 61  
Qy 3794 CGCTGGCTTGGCTCCCAAGTCTGGGATTACAGGTGTA 3835  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 62 CGCTACCTCGGCTCCCAAGTCTGGGATTACAGGTGTA 103

RESULT 5  
LOCUS T77382 103 bp mRNA EST 15-MAR-1995  
DEFINITION Y072h12.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone  
IMAGE:113831 5' similar to contains Alu repetitive element; mRNA  
sequence.  
T77382  
T77382.1 GI:694585  
VERSION  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 103)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT Other-ESTs: Yd72h12.s1  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 943  
Source: IMAGE Consortium, LLNL This clone is available royalty-free  
through LLNL; contact the IMAGE Consortium (info@image.llnl.gov)  
for further information. Putative full length read  
Insert Length: 943 Std Error: 0.00  
Seq primer: M13RP1  
High quality sequence stop: 109.

FEATURES  
source  
1. .103  
/organism="Homo sapiens"  
/db\_xref="GDB:469448"  
/db\_xref="taxon:9606"  
/clone="IMAGE:113831"  
/clone\_lib="Soares fetal liver spleen INFLS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; vector: pT7T3D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;

1st strand cDNA was primed with a Pac I - oligo(dT) primer  
[5' AACTGAGAGATTAATAAGATCTTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 24 a 20 c 37 g 22 t  
ORIGIN

Query Match 0.3%; Score 86.6; DB 21; Length 103;  
Best Local Similarity 95.7%; Pred. No. 0.28;  
Matches 89; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 6073 TGAGAGTCTCACTCTCACTCAACCTCCCTCTCTATATTCAAGTGATCTCTTGCCTCA 6132  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 103 TGAGAGTCTCACTCTCACTCAACCTCCCTCTCTATATTCAAGTGATCTCTTGCCTCA 44  
Qy 6133 GCCTCCCGAGTAGCTGGGACTACAGCGGTGCAC 6165  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 43 GCCTCCCGAGTAGCTGGGACTACAGCGGTGCAC 11

RESULT 6  
LOCUS AA158786/c 106 bp mRNA EST 09-MAR-1998  
DEFINITION z063c11.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone  
IMAGE:591572 5' similar to contains Alu repetitive element; contains  
element P7T7 repetitive element; mRNA sequence.  
AA158786  
AA158786.1 GI:1733588  
VERSION  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 106)  
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.  
WashU-NCI human EST Project  
Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced gi:1406940.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Putative full length read  
The vector to vector length is 119  
Insert Length: 926 Std Error: 0.00  
Seq primer: -28M13 rev2 from Amersham.

FEATURES  
source  
1. .106  
/organism="Homo sapiens"  
/db\_xref="GDB:4622958"  
/db\_xref="taxon:9606"  
/clone="IMAGE:591572"  
/clone\_lib="Stratagene pancreas (#937208)"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: pancreas; Vector: pluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT. Pancreatic adenocarcinoma cell line. Average  
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor  
sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5'  
CTCGAGTCTTTTCTTTTCTTTT 3"  
BASE COUNT 27 a 28 c 37 g 14 t  
ORIGIN

```

Query Match      0.3%; Score 84.8; DB 29; Length 106;
Best Local Similarity 88.5%; Pred. No. 0.45;
Matches 92; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 12168 CTCCTGTTCCAGGCTGGAGTGCATGCGCGCATCTTGCTCAGCAGCAACCTCCGCCTC 12227
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 104 CTCCTGTTCCAGGCTGGAGTGCATGCGCGCATCTTGCTCAGCAGCAACCTCCGCCTC 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 12228 CCGGGTTCAGCATCTCTCTGCTCAGCTCCGGAGTAGCTGG 12271
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 44 CCAAGTTTAAGCAATCTGCTGCCCGCGCTCTCTGAGTGCTGG 1

RESULT 7
A0029690/c
LOCUS      A0029690      109 bp      DNA      GSS      14-APR-1999
DEFINITION RPC111-41F18.TV RPCI-11 Homo sapiens genomic clone RPCI-11-41F18,
genomic survey sequence.
ACCESSION  A0029690
VERSION    A0029690.1 GI:3274821
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 109)
AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
Venter,J.C.
TITLE     Use of BAC End Sequences for Sequence-Ready Map Building (1998)
JOURNAL   Unpublished (1998)
COMMENT   Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdaams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.
FEATURES             Location/Qualifiers
     source            1..109
                     /organism="Homo sapiens"
                     /db_xref="GDB:7515497"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-41F18"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site.1: EcoRI; Site.2: EcoRI;
                     RPC111 Human Male BAC Library"
BASE COUNT          21 a 26 c 24 g 38 t
ORIGIN
Query Match      0.3%; Score 85; DB 94; Length 109;
Best Local Similarity 86.2%; Pred. No. 0.42;
Matches 94; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 4703 CATATCACCTGAGGTGAGGAGTTTGAGACCAGGCTGCCCAACATGGTGAACCCCTGCTC 4762
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 109 CAGATACCTTGAGGTGAGGAGTTTGAGACCAGGCTGCCCAACATGGTGAACCCCTGTATC 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 4763 TACTATAATATATAAAATTAAGCTGGGTGTGGTGTGATGCCCTGTAGTC 4811
||||| || || ||||| ||||| || || || ||||| ||||| || || ||||| ||||| || || |||||

```

---

```

Db 49 TACTAAAACTACAAAATTAGCCGGCATGAAGGAGCATGACTGTATC 1

RESULT 8
A0244245
LOCUS      A0244245      110 bp      mRNA      EST      20-AUG-1997
DEFINITION nc07a04.s1 NCI_CGAP_Prl Homo sapiens cDNA clone IMAGE:1007406
similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION  A0244245
VERSION    A0244245.1 GI:1875104
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 110)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
JOURNAL   On Jan 24, 1995 this sequence version replaced gi:634306.
COMMENT   Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
            M.D., Michael Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LINL at:
            www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -41ml3 fwd. ET from Amersham
High quality sequence stop: 90.
FEATURES             Location/Qualifiers
     source            1..110
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1007406"
                     /clone_lib="NCI_CGAP_Prl"
                     /sex="Male"
                     /dev_stage="45 years old"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; Site.1: NotI; Site.2: EcoRI; 1st
                     strand cDNA was primed with oligo(dT)17 on 50 ng of
                     DNase-treated, total cellular RNA obtained from
                     5,000-10,000 microdissected, histologically normal
                     prostate epithelial cells. Double-stranded cDNA was
                     ligated to EcoRI adaptors, 5 cycles of PCR applied to the
                     cDNA with an adaptor-specific primer, and the resulting
                     PCR product subcloned into pAMP10 by the UDG-cloning
                     method (Life Technologies). Average insert size is 600
                     bp. NOTE: Not directionally cloned. This library was
                     constructed by David Krizman."
BASE COUNT          17 a 26 c 28 g 38 t
ORIGIN
Query Match      0.3%; Score 85; DB 30; Length 110;
Best Local Similarity 85.5%; Pred. No. 0.42;
Matches 94; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 3560 TTTTCTTGTGACACGAGTCTAGCTCTGTGCGCCAGGCTGGAGTGGGACCATCT 3619
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 TTTTCTTGTGAGTGGAGTCTTGTATCTGTGCGCCAGGCTGGAGTGGGACCATCT 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 3620 TGCTCACTCAAGCTCTGCCTCCCGGGTTTATGCCATTCTCATGTCTCA 3669
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 TGGCTCACTCAACCTCTGCCTCTCTGGGTTCAGAGATTCTTCTGCCTCA 110
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 9

AA897366 110 bp mRNA EST 04-JAN-1999  
LOCUS am06h02.s1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
DEFINITION IMAGE:1466067 3' similar to contains Alu repetitive element;; mRNA  
sequence.

AA897366  
ACCESSION AA897366.1 GI:3033986  
KEYWORDS EST.  
SOURCE Homo sapiens

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 110)  
REFERENCE NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2150764.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert length: 834 Std Error: 0.00  
Seq primer: -40ml3 fwd. Et from Amersham  
High quality sequence stop: 63.  
Location/Qualifiers  
1..110  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NDHL19W, testis NHR, and B-cell  
NCI-CGAP-GCBI) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
1.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
BASE COUNT 22 a 27 c 29 g 32 t  
ORIGIN

Query Match 0.3%; Score 85; DB 39; Length 110;  
Best Local Similarity 86.2%; Pred. No. 0.42;  
Matches 94; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 3726 TTTATTTAGTAGAGCGGGTTTCACCGTGTTCAGCCAGAGCGTCTCGATCTCTTGACC 3785  
Db 2 TATTTTATTTAGTAGAGCGGGTTTCACCGTGTTCACCGTGTTCAGCCAGAGCGTCTCTTGACC 61

Qy 3786 TTTCTGATCGCGCTCGCTGGCTTCCAAAGTGCTGGGATTACAGGTGG 3834  
Db 62 TCATGATCGCGCCACCTCGGCTCCCAAGTCTGGGATTATAGCGGTG 110

RESULT 10  
LOCUS AQ535244/c  
DEFINITION RPCI-11-317H22-TV RPCI-11 Homo sapiens genomic clone  
AUTHORS RPCI-11-317H22, genomic survey sequence.  
ACCESSION AQ535244  
VERSION AQ535244.1 GI:4846934  
KEYWORDS GSS.  
SOURCE Homo sapiens

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 103)  
REFERENCE Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and  
Venter,J.C.  
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
Map Building  
JOURNAL Unpublished (1997)  
COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbeetigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieterdejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from  
Research Genet cs ([info@resgen.com](mailto:info@resgen.com)). BAC end search page:  
[http://www.tigr.org/tdb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html).  
Seq primer: T7  
Class: BAC ends  
Location/Qualifiers  
1..103  
/organism="Homo sapiens"  
/db\_xref="GDB:7621533"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-317H22"  
/clone\_lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPCI11 Human Male BAC Library"  
BASE COUNT 31 a 27 c 27 g 18 t  
ORIGIN

Query Match 0.3%; Score 84.4; DB 108; Length 103;  
Best Local Similarity 89.2%; Pred. No. 0.51;  
Matches 91; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 12305 TTTTGTATTTTAGTAGAGCAGGTTTCTCCATGTCGGTCAGGCTGTCTCGAACTCCG 12364  
Db 102 TTTTGTATTTTAGTAGAGCAGGCGGGTTTTCACCATGTTGGCAGGCTGTCTCGAACTCCT 43

Qy 12365 GACATCAGGTGATCTGCCCGCTTGGCCTCCCAAGTCTCGS 12406  
Db 42 GACCTCAAGTGATCTGCCCGCTTGGCCTCCCAAGTCTCGG 1

RESULT 11  
LOCUS AA703692 106 bp mRNA EST 24-DEC-1997  
DEFINITION ag81a10.r1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone  
IMAGE:1140858 5' similar to contains Alu repetitive element;; mRNA  
sequence.

ACCESSION AA703692  
VERSION AA703692.1 GI:2713610  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 106)  
REFERENCE Hallier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,  
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,  
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
TITLE WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397630.  
Contact: Wilson RK  
Washington University School of Medicine



Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT  
ORIGIN

34 a 27 c 32 g 17 t

Query Match 0.3%; Score 84; DB 33; Length 110;  
Best Local Similarity 86.1%; Pred. No. 0.55;  
Matches 93; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 12153 TGAGACGAAGTTCTCTCTTTCCAGGCTGGAGTGAATGGCGCATGTGGCTCATC 12212

Db 110 TGAATGGAGTTTCTCTCTTTCCAGGCTGGAGTGAATGGCTCATGTGGCTCATC 51

Qy 12213 AGCAACCTCCGCTCCCGGTTCAAGCATCTCTCGCTCAGCCTCC 12260

Db 50 CACCACCTCCGCTCCCGGTTCAAGCATCTCTCGCTCAGCCTCC 3

RESULT 14

AA835205

LOCUS AA835205 101 bp mRNA EST 23-FEB-1998  
DEFINITION ak64h01.s1 Barstead pancreas HPLRB1 Homo sapiens cDNA clone  
IMAGE:1412689 3' similar to contains Alu repetitive  
element; contains element KER repetitive element ;, mRNA sequence.

ACCESSION

AA835205

VERSION AA835205.1 GI:2908933

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eucheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 101)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.

WashU-NCI human EST Project

Unpublished (1997)

On Nov 29, 1993 this sequence version replaced gi:636191.

Contact: Wilton RK

Washington University School of Medicine

444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES

source

1..101

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1412689"

/clone.lib="Barstead pancreas HPLRB1"

/sex="female"

/dev\_stage="adult, 34 years"

/lab\_host="DH10B"

/note="Organ: pancreas; Vector: pT73b-Pac (Pharmacia)  
with a modified polylinker; Site.1: EcoRI; Site.2: NotI;  
1st strand cDNA was primed with a Not I - oligo(dT) primer  
[5',

TGTTACGAATGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTTTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
[AATCGATCCTTG], digested with Not I and cloned into the  
Not I and Eco RI sites of the modified pT73 vector.

Library constructed by Bob Barstead."

BASE COUNT

ORIGIN

14 a 36 c 27 g 24 t

Query Match 0.3%; Score 83.4; DB 39; Length 101;  
Best Local Similarity 89.1%; Pred. No. 0.67;  
Matches 90; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 3569 TGAGACGGAGTCTAGCTCTGTGCGCCAGGCTGGAGTGGCAGTGCGCACCACATCTGGCTCACT 3628

Db 1 TGAGACGGAGTCTAGCTCTGTGCGCCAGGCTGGAGTGGCAGTGCTGCTCGGCTCACT 60

Qy 3629 GCAAGCTGCTCCCGGGTTTATGCCAATCTCATGTCTCA 3669

Db 61 GCAAGCTCCGCTCCCGGGTTTACGCCAATCTCTCTGCGCTCA 101

RESULT 15

AI991750

LOCUS AI991750 106 bp mRNA EST 08-SEP-1999

DEFINITION wt48e01.x1 NCI\_CGAP\_Pan1 Homo sapiens cDNA clone IMAGE:2510712 3'  
similar to contains Alu repetitive element; contains element LTR8  
repetitive element ;, mRNA sequence.

ACCESSION

AI991750

VERSION AI991750.1 GI:5838578

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 106)  
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

On Dec 20, 1995 this sequence version replaced gi:1133359.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Life Technologies catalog #: 11548-013

DNA Sequencing by: Washington University Genome Sequencing Center

Clone Distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.blo.llnl.gov/bbrp/image/image.html

FEATURES

source

1..106

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2510712"

/clone.lib="NCI\_CGAP\_Pan1"

/tissue\_type="adenocarcinoma"

/lab\_host="DH10B"

/note="Organ: pancreas; Vector: pCMV-SPORT6; Site.1: SalI;

Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.72 kb. Life Technologies catalog #:

11548-013"

BASE COUNT 24 a 23 c 22 g 37 t

ORIGIN

Query Match 0.3%; Score 83.6; DB 63; Length 106;

Best Local Similarity 86.8%; Pred. No. 0.62;

Matches 92; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 6383 TTGTCTAGTTTATAGAGATGGGGTTTCGCCATCTTGGCCAGGCTGGCTCAAACTCCT 6442

Db 1 TTTTCTTTTAAATAGAGATGAGGTTTACCATGTTGGCAAGGTAGTCTCAAACTCCT 60

Qy 6443 GACCTCAGGTATCTACCCACCTCAGCCTCCCAAGTCTGGGATT 6488

Db 61 GACCTCAGGTATCTACCCACCTCAGCCTCCCAAGTCTGGGATT 106

Search completed: June 17, 2000, 20:31:38  
Job time: 287665 sec

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2000, 17:39:10 ; Search time 372.61 seconds  
(without alignments)  
10116.654 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_1\_29000

Perfect score: 29000

Sequence: 1 CACACACACACACACACA.....AACCTCTGCTCCTGGGTTTC 29000

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 374504

Minimum DB seq length: 10

Maximum DB seq length: 110

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq.\*
- 7: /cgn2\_6/ptodata/1/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	69.8	0.2	105	4	US-08-481-658B-65
c 2	69.6	0.2	105	4	US-08-481-658B-65
3	69.8	0.2	105	4	US-08-477-504A-65
c 4	69.6	0.2	105	4	US-08-477-504A-65
5	69.8	0.2	105	4	US-08-486-756A-65
c 6	69.6	0.2	105	4	US-08-486-756A-65
7	69.8	0.2	105	4	US-08-485-862B-65
c 8	69.6	0.2	105	4	US-08-485-862B-65
9	69.8	0.2	105	5	US-08-787-739-65
c 10	69.6	0.2	105	5	US-08-787-739-65
11	58.6	0.2	78	3	US-08-454-557C-70
c 12	58.6	0.2	78	3	US-08-454-557C-70
13	58.6	0.2	78	4	US-08-450-673C-70
c 14	58.6	0.2	78	4	US-08-450-673C-70
15	58.6	0.2	78	6	PCT-US95-17111A-70
c 16	57.2	0.2	78	3	US-08-454-557C-70
c 17	57.2	0.2	78	4	US-08-340-426D-70
c 18	57.2	0.2	78	4	US-08-450-673C-70
c 19	54.8	0.2	76	3	US-08-454-557C-69
c 20	54.8	0.2	76	4	US-08-340-426D-69
c 21	54.8	0.2	76	4	US-08-450-673C-69
c 22	54.8	0.2	76	6	PCT-US95-17111A-69
c 23	55	0.2	84	3	US-08-454-557C-91
c 24	55	0.2	84	4	US-08-340-426D-91
c 25	55	0.2	84	4	US-08-450-673C-91
c 26	55	0.2	84	6	PCT-US95-17111A-91
27	54.2	0.2	84	3	US-08-454-557C-91

Sequence 91, Appl  
Sequence 91, Appl  
Sequence 91, Appl  
Sequence 60, Appl  
Sequence 60, Appl  
Sequence 60, Appl  
Sequence 60, Appl  
Sequence 60, Appl  
Sequence 92, Appl  
Sequence 92, Appl  
Sequence 92, Appl  
Sequence 36, Appl  
Sequence 57, Appl  
Sequence 57, Appl  
Sequence 57, Appl  
Sequence 60, Appl  
Sequence 60, Appl

ALIGNMENTS

RESULT 1  
US-08-481-658B-65  
; Sequence 65, Application US/08481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,658B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.3E  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHEetical: NO  
; ANTI-SENSE: NO  
US-08-481-658B-65

Query Match 0.2%; Score 69.8; DB 4; Length 105;  
Best Local Similarity 79.0%; Pred. No. 9.6e-07;

Matches	83;	Conservative	0;	Mismatches	22;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

[illegible]

**Qy** 3780 TTGACCTTCTGATCCGGCTGCCTTGCGTTCCCAAAGTGCTGGGAT 3824  
||||| ||||| || |||| |  
**Db** 61 CTGACCTTGTGATCCACGAGCCTCGGCCCTCCCAAAGTGCTGGGAT 105

## RESULT 2

```

US-08-481-658B-65/c
; Sequence 65, Application US/08481658B
; Patent No. 5955075
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jatomir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920

```

Query Match 0.2%; Score 69.6; DB 4; Length 105;  
Best Local Similarity 85.6%; Pred. No. 1.1e-06;  
Matches 89; Conservative 0; Mismatches 14; Indels

QY 1393 ATCTCAGCACATTGGGAGGCTGAGG-GCACAGATCACGAGGTCGGAGTTTGAGACGAGC 1451  
 DB 105 ATCCCAGCACTTTGGGAGGCGGAGGCTGTGGGATCACAGGTCAGGAGTTTGAGAGGAGC 46

QY 1452 CTGGCCAATATGGCGAAACCCCTGTCTCTACTATAAAATACAAAA 1495  
|||||  
Db 45 CTGGCCAATATGGTGAACCCCTGTCTCTACTAAAGATGTA AAAA 2

### RESULTS

```

US-08-477-504A-65
; Sequence 65, Application US/08477504A
; Patent No. 5972353
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,504A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
US-08-477-504A-65

```

Query Match 0.2%; Score 69.8; DB 4; Length 105;  
Best Local Similarity 79.0%; Pred. No. 9.6e-07;  
Matches 83; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy	3720	TTTTTTTTTATTTT	TAGAGATGGGGTTT	CACCTGTTT	AGCGAAGCGTCT	CGATCTC	3779
Db	1	TTTTTTTACATCT	TTTATAGAGACAGGTT	TTCACCATAT	TATGGCCAGGCT	GCTCTCAAACTC	60
Qy	3780	TTGACCTCTTGAT	CCGCCCTGCTCT	TGGCTTCC	CAAAAGTGCTGGGAT		3824
Db	61	CTGACCTTTTGAT	TCCACCAAGCTCG	GGCCCTCC	CAAAAGTGCTGGGAT		105

```

RESULT      4
US -08-477-504A-65/c
; Sequence 65, Application US/08477504A
; Patent No. 5973353
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jatomir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESS: Leona L. Lauder
; STREET: 6 Mariposa Court

```

CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,504A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-477-504A-65

Query Match 0.2%; Score 69.6; DB 4; Length 105;  
Best Local Similarity 85.6%; Pred. No. 1.1e-06;  
Matches 89; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 1393 ATCTCAGCACTTTGGAGGCTGCTGACACATCAGGAGTGGGAGTTTCAGACCAGC 1451  
DB 105 ATCCAGCAGCTTTGGAGGCCGAGGCTGTGTGATCACAAGGTCAGAGTTTGAGAGCAGC 46

QY 1452 CTGGCCAATATGCGCAAAACCTGTCTCTACTATAAATAACAAAA 1495  
DB 45 CTGGCCAATATGCGCAAAACCTGTCTCTACTATAAAGATGTAAAAA 2

RESULT 5  
US-08-486-756A-65  
Sequence 65, Application US/08486756A  
Patent No. 5981711  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,756A  
FILING DATE: 07-JUN-1995

CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3C  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-486-756A-65

Query Match 0.2%; Score 69.8; DB 4; Length 105;  
Best Local Similarity 79.0%; Pred. No. 9.6e-07;  
Matches 83; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 3720 TTTTATTTTATTTTACTAGATGGGGTTTCCACCGTTTACCGAGTGGTTCGATCTC 3779  
DB 1 TTTTATCATCTTTAGTAGACAGGGTTTCCACATATTTGGCCAGGCTGCTCTCAAATC 60

QY 3780 TTGACCTTCTGATCCGCCCTTGGCTTCCCAAAAGTCTGGGAT 3824  
DB 61 CTGACCTTGTGATCCACAGAGCTCGGCTCCCAAAAGTCTGGGAT 105

RESULT 6  
US-08-486-756A-65/c  
Sequence 65, Application US/08486756A  
Patent No. 5981711  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,756A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3C  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-486-756A-65

Query Match      0.2%; Score 69.6; DB 4; Length 105;
Best Local Similarity 85.6%; Pred. No. 1.1e-06;
Matches 89; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 1393 ATCTCAGCACTTTGGAGGCTGAGG-GCACAGATCACGAGTCCGGAGCTTCGAGACCAGC 1451
DB 105 ATCCCAGCACTTTGGAGGCCGAGGCTGGTGATCACAGGTCAGAGGTTTGAGAGCAGC 46

QY 1452 CTGGCAATATGCGAAACCCCTGTCTCTACTAAAAATACAAAA 1495
DB 45 CTGGCAATATGCTGAACCCCTGTCTCTACTAAAGATGTAAAAA 2

RESULT 7
US-08-485-862B-65
; Sequence 65, Application US/08485862B
; Patent No. 5989838
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,862B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/477,504
; FILING DATE: 07-JUN-1995
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-485-862B-65
```

```
Query Match      0.2%; Score 69.8; DB 4; Length 105;
Best Local Similarity 79.0%; Pred. No. 9.6e-07;
Matches 83; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 3720 TTTTATTTTATTTTATAGATAGAGTGGGTTTCCCGTGTAGCCAGAGGTTCTGCATCTC 3779
DB 1 TTTTATCATCTTTATAGTACAGACAGGTTTCCACCATATTTGCCAGGCTGCTCTCAAACTC 60

QY 3780 TTGACCTTCTGATCCGCTTGGCTTCCCAAAAGTCTGGGAT 3824
DB 61 CTGACCTTGTGATCCACGAGCTCGGCTCCCAAAGTCTGGGAT 105

RESULT 8
US-08-485-862B-65/C
; Sequence 65, Application US/08485862B
; Patent No. 5989838
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,862B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/477,504
; FILING DATE: 07-JUN-1995
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-485-862B-65

Query Match      0.2%; Score 69.6; DB 4; Length 105;
Best Local Similarity 85.6%; Pred. No. 1.1e-06;
Matches 89; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 1393 ATCTCAGCACTTTGGAGGCTGAGG-GCACAGATCACGAGTCCGGAGTTCGAGACCAGC 1451
DB 105 ATCCCAGCACTTTGGAGGCCGAGGCTGGTGATCACAGGTCAGAGGTTTGAGAGCAGC 46

QY 1452 CTGGCAATATGCGAAACCCCTGTCTCTACTAAAAATACAAAA 1495
```

US-08-485-862B-65

Db 45 CTGGCAATATGCTGAACCCGTCTCTACTAAAGATGTAAAAA 2

RESULT 9  
US-08-787-739-65  
; Sequence 65, Application US/08787739  
; Patent No. 6027887  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 96  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 369 Pine Street, Suite 610  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/787,739  
; FILING DATE: 24-JAN-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,049  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/486,756  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/477,504  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/481,658  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,862  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,863  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/487,077  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-981-2034  
; TELEFAX: 415-981-0332  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-787-739-65

Query Match 0.2%; Score 69.8; DB 5; Length 105;  
Best Local Similarity 79.0%; Pred. No. 9.6e-07;  
Matches 83; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 3720 TTTTATTTTATTAGTAGAGATGGGTTTCCACCGTGTAGCCAGAACGGTCTCGATCTC 3779  
Db 1 TTTTATTTTATTAGTAGAGATGGGTTTCCACCGTGTAGCCAGAACGGTCTCGATCTC 60  
QY 3780 TTGACCTTCTGATCGCCCTGCGCTTGGCTTCCCAAAAGTGTGGGAT 3824  
Db 61 CTGACCTTGTGATCCACCACCGCTCGGCTCCCAAAAGTGTGGGAT 105

RESULT 10  
US-08-787-739-65/c  
; Sequence 65, Application US/08787739  
; Patent No. 6027887  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 96  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 369 Pine Street, Suite 610  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/787,739  
; FILING DATE: 24-JAN-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,049  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/486,756  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/477,504  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/481,658  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,862  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,863  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/487,077  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-981-2034  
; TELEFAX: 415-981-0332  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-787-739-65

Query Match 0.2%; Score 69.6; DB 5; Length 105;  
Best Local Similarity 85.6%; Pred. No. 1.1e-06;  
Matches 89; Conservative 0; Mismatches 14; Indels 1; Gaps 1;  
Qy 1393 ATCTCAGCAGCTTTGGGAGCTGAGG-GCACAGATCAGAGTCGGGAGTTTGACAGCAGC 1451  
|||||  
Db 105 ATCCAGCAGCTTTGGGAGCGGAGCTGGTGCATCACAAGTTCAGGAGTTTGAGAGCAGC 46  
|||||  
Qy 1452 CTGGCAATATATGGCGAAACCTGTCTCTACTAAATAACAAAA 1495  
|||||  
Db 45 CTGGCAATATATGGTGAACCCCTGTCTCTACTAAAGATCTAAAAA 2

RESULT 11  
US-08-454-557C-70  
; Sequence 70, Application US/08454557C  
; Patent No. 5830670  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; TITLE OF INVENTION: of Alzheimer's Disease  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/454,557C  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, Steven R.  
; REGISTRATION NUMBER: 36,203  
; REFERENCE/DOCKET NUMBER: 0609.3840003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 70:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 78 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
US-08-454-557C-70

Query Match 0.2%; Score 58.6; DB 3; Length 78;  
Best Local Similarity 87.7%; Pred. No. 0.00025;  
Matches 64; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
Qy 2840 GCCCAGCTAAATTTTGTATTTTAGTAGAGATGGGGTTTCTACTATGTTGGCCAGGCTAGT 2899  
|||||  
Db 6 GCCCAGCTAAATTTTGTATTTTAGTAGAGATGGGGTTTCTCTCCATGTTTCATCAGGCTGTT 65  
|||||  
Qy 2900 TTGGAACCTCTGA 2912  
|  
Db 66 GTCGAACCTCTGA 78

RESULT 12  
US-08-340-426D-70  
; Sequence 70, Application US/08340426D  
; Patent No. 5948634

; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; APPLICANT: Wands, Jack R.  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; TITLE OF INVENTION: of Alzheimer's Disease  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/340,426D  
; FILING DATE: 14-NOV-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, Steven R.  
; REGISTRATION NUMBER: 36,203  
; REFERENCE/DOCKET NUMBER: 0609.3840002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 70:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 78 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
US-08-340-426D-70

Query Match 0.2%; Score 58.6; DB 4; Length 78;  
Best Local Similarity 87.7%; Pred. No. 0.00025;  
Matches 64; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
Qy 2840 GCCCAGCTAAATTTTGTATTTTAGTAGAGATGGGGTTTCTACTATGTTGGCCAGGCTAGT 2899  
|||||  
Db 6 GCCCAGCTAAATTTTGTATTTTAGTAGAGATGGGGTTTCTCTCCATGTTTCATCAGGCTGTT 65  
|||||  
Qy 2900 TTGGAACCTCTGA 2912  
|  
Db 66 GTCGAACCTCTGA 78

RESULT 13  
US-08-450-673C-70  
; Sequence 70, Application US/08450673C  
; Patent No. 5948888  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; APPLICANT: Wands, Jack R.  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; TITLE OF INVENTION: of Alzheimer's Disease  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,673C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
;
US-08-450-673C-70

Query Match 0.2%; Score 58.6; DB 4; Length 78;
Best Local Similarity 87.7%; Pred. No. 0.00025;
Matches 64; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2840 GCCCAGCTAATTTTGTATTTTCTAGATGGGTTTCTACTATGTTGCCAGGCTAGT 2899
Db 6 GCCCAGCTAATTTTGTATTTTCTAGATGGGTTTCTCTCATGTTCCATCAGGCTGCT 65

QY 2900 TTGGAACCTCTGA 2912
Db 66 GTCGAACCTCTGA 78

RESULT 14
PCT-US95-17111A-70
; Sequence 70, Application PC/TUS9517111A
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and
; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 131
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
;
US-08-450-673C-70

Query Match 0.2%; Score 58.6; DB 4; Length 78;
Best Local Similarity 87.7%; Pred. No. 0.00025;
Matches 64; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2840 GCCCAGCTAATTTTGTATTTTCTAGATGGGTTTCTACTATGTTGCCAGGCTAGT 2899
Db 6 GCCCAGCTAATTTTGTATTTTCTAGATGGGTTTCTCTCATGTTCCATCAGGCTGCT 65

QY 2900 TTGGAACCTCTGA 2912
Db 66 GTCGAACCTCTGA 78

RESULT 14
PCT-US95-17111A-70
; Sequence 70, Application PC/TUS9517111A
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and
; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 131
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
;
US-08-450-673C-70
```

```
;
; TOPOLOGY: both
;
PCT-US95-17111A-70

Query Match 0.2%; Score 58.6; DB 6; Length 78;
Best Local Similarity 87.7%; Pred. No. 0.00025;
Matches 64; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2840 GCCCAGCTAATTTTGTATTTTCTAGATGGGTTTCTACTATGTTGCCAGGCTAGT 2899
Db 6 GCCCAGCTAATTTTGTATTTTCTAGATGGGTTTCTCTCATGTTCCATCAGGCTGCT 65

QY 2900 TTGGAACCTCTGA 2912
Db 66 GTCGAACCTCTGA 78

RESULT 15
US-08-454-557C-70/C
; Sequence 70, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,557C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
;
US-08-454-557C-70

Query Match 0.2%; Score 57.2; DB 3; Length 78;
Best Local Similarity 83.3%; Pred. No. 0.00051;
Matches 65; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1433 TCGGGAGTTTGAGACCAGCTGCGCCCAATATGCGGAACCCCTGCTCTCTACTAAATAACAA 1492
Db 78 TCAGGAGTTTCGACACCAGCCTGATGAACATGGAGAAACCCCATCTCTACTAAATAACAA 19

QY 1493 AAATTAGCTGGCGCTGCT 1510
Db 18 ATATTAGCTGGCGCTGCT 1
```

Search completed: June 18, 2000, 01:37:37  
Job time: 304987 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2000, 01:31:16 ; Search time 17971.2 Seconds  
(without alignments)  
-1569.844 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_28000\_57000

Perfect score: 29001

Sequence: 1 GATAATTATTTTAATATAT.....ACCCCTTTGATGAGAACATGA 29001

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 370290

Minimum DB seq length: 10

Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_p11.\*
- 8: gb\_p12.\*
- 9: gb\_p1.\*
- 10: gb\_p12.\*
- 11: gb\_p13.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_vi.\*
- 17: em\_fun.\*
- 18: em\_humi.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_om.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pl.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_vi.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_ba1.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*
- 44: gb\_htg6.\*

- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*
- 50: gb\_p13.\*
- 51: gb\_pr5.\*
- 52: gb\_htg8.\*
- 53: gb\_htg9.\*
- 54: gb\_htg10.\*
- 55: gb\_htg11.\*
- 56: gb\_htg12.\*
- 57: gb\_htg13.\*
- 58: gb\_htg14.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	94.4	0.3	96	13	G31304	G31304 sy899g1-19
2	91.6	0.3	108	10	HSJDLRN2	X05250 Human LDL-r
3	88.8	0.3	108	10	HSJDLRN2	X05250 Human LDL-r
4	88	0.3	108	11	HSU67803	U67803 Human LDL-r
5	87	0.3	107	9	HUMALCE162	M87924 Human carci
6	87	0.3	108	11	HSU67804	U67804 Human small
7	85.4	0.3	103	9	HUMALCE221	M87896 Human carci
8	84	0.3	103	9	HUMALCE221	M87896 Human carci
9	84.2	0.3	108	10	HSJDLRD1	X05249 Human LDL-r
10	84.2	0.3	108	10	HSJDLRD2	X05251 Human LDL-r
11	81.4	0.3	108	10	HSJDLI12	X05248 Human LDL-r
12	81.4	0.3	108	10	HSJDLRD1	X05249 Human LDL-r
13	81.4	0.3	108	10	HSJDLRD2	X05251 Human LDL-r
14	79	0.3	108	11	HSU67808	U67808 Human small
15	78.8	0.3	110	9	HUMALCE43	M87900 Human carci
16	78.2	0.3	110	11	HSU67807	U67807 Human small
17	77.4	0.3	90	9	HUMDLRFL	X03555 Human low d
18	77.6	0.3	103	13	HS8IC8R	X57789 Human sequ
19	77.6	0.3	104	9	HUMALCE272	M87899 Human carci
20	77.4	0.3	107	9	HUMALCE162	M87924 Human carci
21	76.2	0.3	108	11	HSU67803	U67803 Human small
22	75.4	0.3	97	9	HUMDLRD3	M14179 Human famil
23	75.6	0.3	103	13	HS8IC8R	X57789 Human sequ
24	75.6	0.3	108	10	HSJDLI12	X05248 Human LDL-r
25	75	0.3	110	11	HSU67807	U67807 Human small
26	73.8	0.3	97	9	HUMDLRA1	M14178 Human low d
27	73.6	0.3	97	9	HUMDLRA2	M14180 Human low d
28	73	0.3	107	11	HSU67806	U67806 Human small
29	73.2	0.3	110	9	HUMALCE43	M87900 Human carci
30	72.6	0.3	100	9	HUMGALNSA	D45223 Human GALNS
31	72.6	0.3	108	9	HUMD1D03M5	D16965 Human HepG2
32	72.2	0.2	108	9	HUMD1D03M5	D16965 Human HepG2
33	71	0.2	104	9	HUMALCE272	M87899 Human carci
34	71	0.2	107	11	HSU67806	U67806 Human small
35	70.6	0.2	101	10	S79560	S79560 HRX {intron
36	69.8	0.2	108	11	HSU67808	U67808 Human small
37	69.2	0.2	91	13	HUMUT8164A	L30244 Human STS U
38	69.2	0.2	99	13	HUMUT7692A	L30306 Human STS U
39	69.2	0.2	100	13	HUMUT931A	L31299 Human STS U
40	69.4	0.2	108	13	G43535	G43535 WIAF-2393-S
41	68.8	0.2	80	9	HUMERKFAE	M36135 Human alpha
42	68.8	0.2	108	13	G32614	G32614 A009K21 Hum
43	67.6	0.2	91	13	HUMUT8164A	L30244 Human STS U
44	67.8	0.2	100	9	HUMGALNSA	D45223 Human GALNS
45	67.6	0.2	107	13	G32919	G32919 A009W27 Hum

ALIGNMENTS

```

RESULT 1
G31304
LOCUS G31304 96 bp DNA STS 29-SEP-1998
DEFINITION sy899g1-19 Human (A.Gnirke) Homo sapiens STS genomic, sequence
tagged site.
ACCESSION G31304
VERSION G31304.1 GI:1871333
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 96)
AUTHORS Lauer,P., Meyer,N.C., Prass,C.E., Starnes,S.M., Wolff,R.K. and
Gnirke,A.
TITLE Clone-ontig and STS maps of the hereditary hemochromatosis region
on human chromosome 6p21.3-6p22
JOURNAL on human chromosome 6p21.3-6p22
MEDLINE Genome Res. 7 (5), 457-470 (1997)
COMMENT 97294058
GDB: GDB:5584195
GDB_DSEG: D6S2377
Contact: Andreas Gnirke
Mercator Genetics, Inc.
4040 Campbell Ave, Menlo Park CA, 94025, USA
Email: gnirke@mercator.com
Primer A: GTCCCAAGCAATATAATGAG
Primer B: AGGCACAGTGGGAG
STS size: 77
PCR Profile:
Denaturation: 92 degrees C for 20 seconds
Annealing: 60 degrees C for 45 seconds
Polymerization: 72 degrees C for 60 seconds
PCR Cycles: 35
Thermal Cycler: MJ Research PTC-200
Protocol:
Template: 30-200 ng
Primer: each 0.8 uM
dNTPs: each 200 uM
Taq Polymerase: 0.05 units/ul
Total Vol: 12 ul
Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3
gelatin: 0.001% (w/v).
Location/Qualifiers
1..96
/organism="Homo sapiens"
/clone_lib="Human (A.Gnirke)"
9..85
primer_bind 9..30
primer_bind complement(70..85)
BASE COUNT 22 a 29 c 23 g 22 t
ORIGIN
STS
primer_bind
primer_bind
complement(70..85)
22 a 29 c 23 g 22 t
Query Match 0.3%; Score 94.4; DB 13; Length 96;
Best Local Similarity 99.0%; Pred. No. 4.2e-08;
Matches 95; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 12823 CTGCTGGTGTCCCAAGCAATATAATGAGAAAATGCTTCCATGGATGCCAGATCCCC 12882
|||||
Db 1 CTGCTGGTGTCCCAAGCAATATAATGAGAAAATGCTTCCATGGATGCCAGATCCCC 60
QY 12883 TCTGCCCCCTCTTCCACATGTGCGCTGGGGCAGAGGT 12918
|||||
Db 61 TCTGCCCCCTCTTCCACATGTGCGCTGGGGCAGAGGT 96

```

```

RESULT 2
HSLDLRN2/c
LOCUS HSLDLRN2 108 bp DNA PRI 20-MAY-1992
DEFINITION Human LDL-receptor gene intron 14 fragment (normal gene).
ACCESSION X05250
VERSION X05250.1 GI:34337
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
Williamson,R., and Humphries,S.
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT See X05252 for deletion junction
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES
source
Location/Qualifiers
1..108
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..108
/note="intron XIV fragment"
BASE COUNT 28 a 23 c 39 g 18 t
ORIGIN
Query Match 0.3%; Score 91.6; DB 10; Length 108;
Best Local Similarity 91.5%; Pred. No. 1.4e-07;
Matches 97; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 10742 CTCGGCTCACTGCAAGCTCTGCCTCTGGTTTCATGCCATTCTCTGCTCAGCCTCCG 10801
|||||
Db 108 CTCGGCTCACTGCAAGCTCTGCCTCTGGTTTCATGCCATTCTCTGCTCAGCCTCCG 49
QY 10802 AGTAGCTGGGACTACAGCGTCTGCCACCACGCCAGCTAATTTTT 10847
|||||
Db 48 AGTAGCTGGGATTACAGGCACCTGCCACCACGCCGTGCTAATTTTT 3
RESULT 3
HSLDLRN2
LOCUS HSLDLRN2 108 bp DNA PRI 20-MAY-1992
DEFINITION Human LDL-receptor gene intron 14 fragment (normal gene).
ACCESSION X05250
VERSION X05250.1 GI:34337
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
Williamson,R., and Humphries,S.
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT See X05252 for deletion junction
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES
source
Location/Qualifiers
1..108
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..108
/note="intron XIV fragment"
BASE COUNT 28 a 23 c 39 g 18 t
ORIGIN

```

## ORIGIN

Query Match 0.3%; Score 88.8; DB 10; Length 108;  
Best Local Similarity 88.9%; Pred. No. 4.7e-07;  
Matches 96; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 7126 AAAAACTTAGCGTCATGGTGGCCGCTGTAGTCTTCAGCCACTTGGGAGGCTGAG 7185

Db 1 ACAAAATTAGCCAGCGTGGTGGCAGGTCCCTGTATCCCACTACTCGGAGGCTGAG 60

QY 7186 GCAGGAAATTCGTTGAACCCAGGAGCGAGGTTGCAGTGAGCCCGAG 7233

Db 61 GCAGGAAATTCGTTGAACCCAGGAGCGAGGTTGCAGTGAGCCCGAG 108

## RESULT 4

HSU67803/c 108 bp RNA PRI 01-AUG-1997

LOCUS Human small cytoplasmic Alu transcript.

DEFINITION U67803

ACCESSION U67803.1 GI:2289917

VERSION Alu.

KEYWORDS human.

SOURCE

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 108)

AUTHORS Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.

TITLE cDNAs derived from primary and small cytoplasmic Alu (scAlu)

JOURNAL J. Mol. Biol. 271 (2), 222-234 (1997)

MEDLINE 97415756

AUTHORS 2 (bases 1 to 108)

Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.

TITLE Direct Submission

JOURNAL Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The

Children's Hospital of Philadelphia, 1004F Abramson Research

Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA

FEATURES

source

1. .108

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="TscAlu2"

repeat\_region 1. .108

/note="scAlu"

/rpt\_family="Alu"

/rpt\_type-dispersed

BASE COUNT 23 a 39 c 30 g 16 t

## ORIGIN

Query Match 0.3%; Score 88; DB 11; Length 108;

Best Local Similarity 94.8%; Pred. No. 6.7e-07;

Matches 91; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 10859 TAGAGATGGGGTTTACCATTAGCCAGGATGCTTCGATCTCCGACCTCGTGATCCA 10918

Db 96 TAGACAGGGGGTTTACCTTGTGTAGCCAGGATGCTTCGATCTCCGACCTCGTGATCCG 37

QY 10919 CCGCTTTTGGCCCTCCCAAAGTGTGGATTACAGGC 10954

Db 36 CCGCTTGGCCCTCCCAAAGTGTGGATTACAGGC 1

## RESULT 5

HUMALCE162/c

LOCUS HUMALCE162 107 bp ss-RNA PRI 15-APR-1994

DEFINITION Human carcinoma cell-derived Alu RNA transcript, clone CE162.

ACCESSION M87924

VERSION M87924.1 GI:174871

KEYWORDS Alu repeat.

SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 107)

AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.

TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of

post-transcriptional selection of master sequences

JOURNAL J. Mol. Biol. (1992) In press

FEATURES

source

1. .107

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/cell\_line="Ntera2D1"

/dev\_stage="embryo"

/sex="male"

/tissue\_type="carcinoma"

BASE COUNT 28 a 30 c 35 g 14 t

## ORIGIN

Query Match 0.3%; Score 87; DB 9; Length 107;

Best Local Similarity 90.3%; Pred. No. 1e-06;

Matches 93; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 10688 TTTTGTGAGATGGAGTCTCACTCTGTACCTAGGCTGGAGTGGCGCAAACTCGC 10747

Db 107 TTTTGTGAGAGGAGTCTCGTCTGTGCCCCAGGCTGGAGTGGCGGATCTCGC 48

QY 10748 TCACTGCAAGCTTGCCTCTCTGGTTCATGCCATTCCTCTGCC 10790

Db 47 TCACTGCAAGCTCGGCTCCCGGGTTCACGCCATTCCTCTCTGCC 5

## RESULT 6

HSU67804/c

LOCUS HSU67804 108 bp RNA PRI 01-AUG-1997

DEFINITION Human small cytoplasmic Alu transcript.

ACCESSION U67804

VERSION U67804.1 GI:2289918

KEYWORDS Alu.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 108)

AUTHORS Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.

TITLE cDNAs derived from primary and small cytoplasmic Alu (scAlu)

JOURNAL J. Mol. Biol. 271 (2), 222-234 (1997)

MEDLINE 97415756

AUTHORS 2 (bases 1 to 108)

Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.

TITLE Direct Submission

JOURNAL Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The

Children's Hospital of Philadelphia, 1004F Abramson Research

Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA

FEATURES

source

1. .108

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="TscAlu3"

repeat\_region 1. .108

/note="scAlu"

/rpt\_family="Alu"

/rpt\_type-dispersed

BASE COUNT 26 a 38 c 26 g 18 t

## ORIGIN

Query Match 0.3%; Score 87; DB 11; Length 108;

Best Local Similarity 94.7%; Pred. No. 1e-06;

Matches 90; Conservative 0; Mismatches 5; Indels 0; Gaps 0;



```

VERSION X05251.1 GI:34336
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
          Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
          Williamson,R. and Humphries,S.
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT *source: hypercholesterol aemia
See X05250 for corresponding normal gene sequence
In the defective LDL-receptor gene the deletion occurred between two
alu-repetitive sequences, that are in the same direction, the
deletion eliminates exons 13 and 14 and changes the reading frame
of the resulting spliced mRNA.
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES
  source      1..108
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
  Intron
  BASE COUNT 28 a 20 c 40 g 20 t
  ORIGIN
          1..108
          /note="intron XIV fragment"

  Query Match      0.3%; Score 84.2; DB 10; Length 108;
  Best Local Similarity 87.6%; Pred. No. 3 5e-06;
  Matches 92; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 10743 TCGGCTCACTGCAAGCTCTCCCTCCCTGGGTTCAATTCCTCGCTCAGCCTCCCGA 10802
          ||| |||| ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 107 TCGCCTCACCACACCTCTCCCTCCCTGGGTTCAACCACTTTCTCGCTCAGCCTCCCGA 48

QY 10803 GTAGCTGGGATTACAGCGCTCGCCACCCAGCCCGCAGCTAATTTT 10847
          ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 47 GTAGCTGGGATTACAGCGCTCGCCACCCAGCCCGCAGCTAATTTT 3

RESULT 11
HSLDL112
LOCUS Human LDL-receptor gene intron 12 fragment (normal gene) LDL = low
DEFINITION density lipoprotein.
ACCESSION X05248
VERSION X05248.1 GI:34334
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor;
          repetitive sequence.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
          Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
          Williamson,R. and Humphries,S.
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT See X05249 for deletion junction
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES
  source      1..108
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
  misc_feature 20 a 20 c 40 g 20 t
  BASE COUNT 28 a 20 c 40 g 20 t
  ORIGIN
          1..108
          /note="intron XIV fragment"

```

```

          /note="Alu repeat"
          1..108
          /note="intron XII fragment"
  BASE COUNT 21 a 38 c 20 g 29 t
  ORIGIN
          1..108
          /note="Alu repeat"

  Query Match      0.3%; Score 81.4; DB 10; Length 108;
  Best Local Similarity 85.0%; Pred. No. 1.2e-05;
  Matches 91; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 5414 TCGGCTCACCACCACTCTACCTCCAGGTTCAAGCAATTCCTCGCTCAGCCTCCCGA 5473
          ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 2 TCGCCTCACCACCACTCTCGCTCCCTGGGTTCAAAACCACTTTCTCGCTCAGCCTCCTTA 61

QY 5474 GTAGCTGGGATTACAGCGCTGCATCACCACCCAGCTAATTTTGTGA 5520
          ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 62 GTAGCTGGGATTACAGCATGTGCCACCCAGCCCGGCTGATTTGTGA 108

RESULT 12
HSLDLRD1/c
LOCUS Human LDL-receptor mutated gene with intron 12 deletion junction.
DEFINITION X05249
ACCESSION X05249.1 GI:34335
VERSION X05249.1
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
          Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
          Williamson,R. and Humphries,S.
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT *source: hypercholesterol aemia
See X05248 for corresponding normal gene sequence
In the defective LDL-receptor gene the deletion occurred between two
alu-repetitive sequences, that are in the same direction, the
deletion eliminates exons 13 and 14 and changes the reading frame
of the resulting spliced mRNA.
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES
  source      1..108
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
  misc_feature 1..108
              /note="deletion junction region intron 12/ intron 15"
  BASE COUNT 20 a 40 c 20 g 28 t
  ORIGIN
          1..108
          /note="deletion junction region intron 12/ intron 15"

  Query Match      0.3%; Score 81.4; DB 10; Length 108;
  Best Local Similarity 85.0%; Pred. No. 1.2e-05;
  Matches 91; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 7126 AAAAATCTTAGCGCTGATGCTGCATGCCCTGTACTCTCAGCACTTGGAGGCTGAG 7185
          ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 108 AAAAAATTTAGCAGCGCTGCTGCAGGTGCTGTATATCCAGCTACTCCGGAGGCTGAG 49

QY 7186 GCAGGAAATTTGCTTGAACCCAGGAGGAGGTTGAGTTCAGTGAGCCGA 7232
          ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 48 GCAGGAAATTTGTTGAACCCAGGAGGAGGTTGTTGGTGAGGCCGA 2

RESULT 13
HSLDLRD2
LOCUS
          HSLDLRD2      108 bp      DNA      PRI      20-MAY-1992

```

```

DEFINITION Human LDL-receptor mutated gene with intron 14 deletion junction.
ACCESSION X05251
VERSION X05251.1 GI:34336
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 108)
Horshtemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
Williamson,R. and Humphries,S.
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT *source: hypercholesterol aemia
See X05250 for corresponding normal gene sequence
In the defective LDL-receptor gene the deletion occurred between two
alu-repetitive sequences, that are in the same direction, the
deletion eliminates exons 13 and 14 and changes the reading frame
of the resulting spliced mRNA.
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES
    source
        1..108
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /cell_type="blood leukocytes from a patient with familial"
    intron
        1..108
        /note="intron XIV fragment"
BASE COUNT 28 a 20 c 40 g 20 t
ORIGIN
    1
    2
    3
    4
    5
    6
    7
    8
    9
    10
    11
    12
    13
    14
    15
    16
    17
    18
    19
    20
    21
    22
    23
    24
    25
    26
    27
    28
    29
    30
    31
    32
    33
    34
    35
    36
    37
    38
    39
    40
    41
    42
    43
    44
    45
    46
    47
    48
    49
    50
    51
    52
    53
    54
    55
    56
    57
    58
    59
    60
    61
    62
    63
    64
    65
    66
    67
    68
    69
    70
    71
    72
    73
    74
    75
    76
    77
    78
    79
    80
    81
    82
    83
    84
    85
    86
    87
    88
    89
    90
    91
    92
    93
    94
    95
    96
    97
    98
    99
    100
    101
    102
    103
    104
    105
    106
    107
    108
    109
    110
    111
    112
    113
    114
    115
    116
    117
    118
    119
    120
    121
    122
    123
    124
    125
    126
    127
    128
    129
    130
    131
    132
    133
    134
    135
    136
    137
    138
    139
    140
    141
    142
    143
    144
    145
    146
    147
    148
    149
    150
    151
    152
    153
    154
    155
    156
    157
    158
    159
    160
    161
    162
    163
    164
    165
    166
    167
    168
    169
    170
    171
    172
    173
    174
    175
    176
    177
    178
    179
    180
    181
    182
    183
    184
    185
    186
    187
    188
    189
    190
    191
    192
    193
    194
    195
    196
    197
    198
    199
    200
    201
    202
    203
    204
    205
    206
    207
    208
    209
    210
    211
    212
    213
    214
    215
    216
    217
    218
    219
    220
    221
    222
    223
    224
    225
    226
    227
    228
    229
    230
    231
    232
    233
    234
    235
    236
    237
    238
    239
    240
    241
    242
    243
    244
    245
    246
    247
    248
    249
    250
    251
    252
    253
    254
    255
    256
    257
    258
    259
    260
    261
    262
    263
    264
    265
    266
    267
    268
    269
    270
    271
    272
    273
    274
    275
    276
    277
    278
    279
    280
    281
    282
    283
    284
    285
    286
    287
    288
    289
    290
    291
    292
    293
    294
    295
    296
    297
    298
    299
    300
    301
    302
    303
    304
    305
    306
    307
    308
    309
    310
    311
    312
    313
    314
    315
    316
    317
    318
    319
    320
    321
    322
    323
    324
    325
    326
    327
    328
    329
    330
    331
    332
    333
    334
    335
    336
    337
    338
    339
    340
    341
    342
    343
    344
    345
    346
    347
    348
    349
    350
    351
    352
    353
    354
    355
    356
    357
    358
    359
    360
    361
    362
    363
    364
    365
    366
    367
    368
    369
    370
    371
    372
    373
    374
    375
    376
    377
    378
    379
    380
    381
    382
    383
    384
    385
    386
    387
    388
    389
    390
    391
    392
    393
    394
    395
    396
    397
    398
    399
    400
    401
    402
    403
    404
    405
    406
    407
    408
    409
    410
    411
    412
    413
    414
    415
    416
    417
    418
    419
    420
    421
    422
    423
    424
    425
    426
    427
    428
    429
    430
    431
    432
    433
    434
    435
    436
    437
    438
    439
    440
    441
    442
    443
    444
    445
    446
    447
    448
    449
    450
    451
    452
    453
    454
    455
    456
    457
    458
    459
    460
    461
    462
    463
    464
    465
    466
    467
    468
    469
    470
    471
    472
    473
    474
    475
    476
    477
    478
    479
    480
    481
    482
    483
    484
    485
    486
    487
    488
    489
    490
    491
    492
    493
    494
    495
    496
    497
    498
    499
    500
    501
    502
    503
    504
    505
    506
    507
    508
    509
    510
    511
    512
    513
    514
    515
    516
    517
    518
    519
    520
    521
    522
    523
    524
    525
    526
    527
    528
    529
    530
    531
    532
    533
    534
    535
    536
    537
    538
    539
    540
    541
    542
    543
    544
    545
    546
    547
    548
    549
    550
    551
    552
    553
    554
    555
    556
    557
    558
    559
    560
    561
    562
    563
    564
    565
    566
    567
    568
    569
    570
    571
    572
    573
    574
    575
    576
    577
    578
    579
    580
    581
    582
    583
    584
    585
    586
    587
    588
    589
    590
    591
    592
    593
    594
    595
    596
    597
    598
    599
    600
    601
    602
    603
    604
    605
    606
    607
    608
    609
    610
    611
    612
    613
    614
    615
    616
    617
    618
    619
    620
    621
    622
    623
    624
    625
    626
    627
    628
    629
    630
    631
    632
    633
    634
    635
    636
    637
    638
    639
    640
    641
    642
    643
    644
    645
    646
    647
    648
    649
    650
    651
    652
    653
    654
    655
    656
    657
    658
    659
    660
    661
    662
    663
    664
    665
    666
    667
    668
    669
    670
    671
    672
    673
    674
    675
    676
    677
    678
    679
    680
    681
    682
    683
    684
    685
    686
    687
    688
    689
    690
    691
    692
    693
    694
    695
    696
    697
    698
    699
    700
    701
    702
    703
    704
    705
    706
    707
    708
    709
    710
    711
    712
    713
    714
    715
    716
    717
    718
    719
    720
    721
    722
    723
    724
    725
    726
    727
    728
    729
    730
    731
    732
    733
    734
    735
    736
    737
    738
    739
    740
    741
    742
    743
    744
    745
    746
    747
    748
    749
    750
    751
    752
    753
    754
    755
    756
    757
    758
    759
    760
    761
    762
    763
    764
    765
    766
    767
    768
    769
    770
    771
    772
    773
    774
    775
    776
    777
    778
    779
    780
    781
    782
    783
    784
    785
    786
    787
    788
    789
    790
    791
    792
    793
    794
    795
    796
    797
    798
    799
    800
    801
    802
    803
    804
    805
    806
    807
    808
    809
    810
    811
    812
    813
    814
    815
    816
    817
    818
    819
    820
    821
    822
    823
    824
    825
    826
    827
    828
    829
    830
    831
    832
    833
    834
    835
    836
    837
    838
    839
    840
    841
    842
    843
    844
    845
    846
    847
    848
    849
    850
    851
    852
    853
    854
    855
    856
    857
    858
    859
    860
    861
    862
    863
    864
    865
    866
    867
    868
    869
    870
    871
    872
    873
    874
    875
    876
    877
    878
    879
    880
    881
    882
    883
    884
    885
    886
    887
    888
    889
    890
    891
    892
    893
    894
    895
    896
    897
    898
    899
    900
    901
    902
    903
    904
    905
    906
    907
    908
    909
    910
    911
    912
    913
    914
    915
    916
    917
    918
    919
    920
    921
    922
    923
    924
    925
    926
    927
    928
    929
    930
    931
    932
    933
    934
    935
    936
    937
    938
    939
    940
    941
    942
    943
    944
    945
    946
    947
    948
    949
    950
    951
    952
    953
    954
    955
    956
    957
    958
    959
    960
    961
    962
    963
    964
    965
    966
    967
    968
    969
    970
    971
    972
    973
    974
    975
    976
    977
    978
    979
    980
    981
    982
    983
    984
    985
    986
    987
    988
    989
    990
    991
    992
    993
    994
    995
    996
    997
    998
    999
    1000

```

```

repeat_region
    1..108
    /note="scalu"
    /rpt_family="Alu"
    /rpt_type="dispersed"
BASE COUNT 22 a 28 g 21 t
ORIGIN
    1
    2
    3
    4
    5
    6
    7
    8
    9
    10
    11
    12
    13
    14
    15
    16
    17
    18
    19
    20
    21
    22
    23
    24
    25
    26
    27
    28
    29
    30
    31
    32
    33
    34
    35
    36
    37
    38
    39
    40
    41
    42
    43
    44
    45
    46
    47
    48
    49
    50
    51
    52
    53
    54
    55
    56
    57
    58
    59
    60
    61
    62
    63
    64
    65
    66
    67
    68
    69
    70
    71
    72
    73
    74
    75
    76
    77
    78
    79
    80
    81
    82
    83
    84
    85
    86
    87
    88
    89
    90
    91
    92
    93
    94
    95
    96
    97
    98
    99
    100
    101
    102
    103
    104
    105
    106
    107
    108
    109
    110
    111
    112
    113
    114
    115
    116
    117
    118
    119
    120
    121
    122
    123
    124
    125
    126
    127
    128
    129
    130
    131
    132
    133
    134
    135
    136
    137
    138
    139
    140
    141
    142
    143
    144
    145
    146
    147
    148
    149
    150
    151
    152
    153
    154
    155
    156
    157
    158
    159
    160
    161
    162
    163
    164
    165
    166
    167
    168
    169
    170
    171
    172
    173
    174
    175
    176
    177
    178
    179
    180
    181
    182
    183
    184
    185
    186
    187
    188
    189
    190
    191
    192
    193
    194
    195
    196
    197
    198
    199
    200
    201
    202
    203
    204
    205
    206
    207
    208
    209
    210
    211
    212
    213
    214
    215
    216
    217
    218
    219
    220
    221
    222
    223
    224
    225
    226
    227
    228
    229
    230
    231
    232
    233
    234
    235
    236
    237
    238
    239
    240
    241
    242
    243
    244
    245
    246
    247
    248
    249
    250
    251
    252
    253
    254
    255
    256
    257
    258
    259
    260
    261
    262
    263
    264
    265
    266
    267
    268
    269
    270
    271
    272
    273
    274
    275
    276
    277
    278
    279
    280
    281
    282
    283
    284
    285
    286
    287
    288
    289
    290
    291
    292
    293
    294
    295
    296
    297
    298
    299
    300
    301
    302
    303
    304
    305
    306
    307
    308
    309
    310
    311
    312
    313
    314
    315
    316
    317
    318
    319
    320
    321
    322
    323
    324
    325
    326
    327
    328
    329
    330
    331
    332
    333
    334
    335
    336
    337
    338
    339
    340
    341
    342
    343
    344
    345
    346
    347
    348
    349
    350
    351
    352
    353
    354
    355
    356
    357
    358
    359
    360
    361
    362
    363
    364
    365
    366
    367
    368
    369
    370
    371
    372
    373
    374
    375
    376
    377
    378
    379
    380
    381
    382
    383
    384
    385
    386
    387
    388
    389
    390
    391
    392
    393
    394
    395
    396
    397
    398
    399
    400
    401
    402
    403
    404
    405
    406
    407
    408
    409
    410
    411
    412
    413
    414
    415
    416
    417
    418
    419
    420
    421
    422
    423
    424
    425
    426
    427
    428
    429
    430
    431
    432
    433
    434
    435
    436
    437
    438
    439
    440
    441
    442
    443
    444
    445
    446
    447
    448
    449
    450
    451
    452
    453
    454
    455
    456
    457
    458
    459
    460
    461
    462
    463
    464
    465
    466
    467
    468
    469
    470
    471
    472
    473
    474
    475
    476
    477
    478
    479
    480
    481
    482
    483
    484
    485
    486
    487
    488
    489
    490
    491
    492
    493
    494
    495
    496
    497
    498
    499
    500
    501
    502
    503
    504
    505
    506
    507
    508
    509
    510
    511
    512
    513
    514
    515
    516
    517
    518
    519
    520
    521
    522
    523
    524
    525
    526
    527
    528
    529
    530
    531
    532
    533
    534
    535
    536
    537
    538
    539
    540
    541
    542
    543
    544
    545
    546
    547
    548
    549
    550
    551
    552
    553
    554
    555
    556
    557
    558
    559
    560
    561
    562
    563
    564
    565
    566
    567
    568
    569
    570
    571
    572
    573
    574
    575
    576
    577
    578
    579
    580
    581
    582
    583
    584
    585
    586
    587
    588
    589
    590
    591
    592
    593
    594
    595
    596
    597
    598
    599
    600
    601
    602
    603
    604
    605
    606
    607
    608
    609
    610
    611
    612
    613
    614
    615
    616
    617
    618
    619
    620
    621
    622
    623
    624
    625
    626
    627
    628
    629
    630
    631
    632
    633
    634
    635
    636
    637
    638
    639
    640
    641
    642
    643
    644
    645
    646
    647
    648
    649
    650
    651
    652
    653
    654
    655
    656
    657
    658
    659
    660
    661
    662
    663
    664
    665
    666
    667
    668
    669
    670
    671
    672
    673
    674
    675
    676
    677
    678
    679
    680
    681
    682
    683
    684
    685
    686
    687
    688
    689
    690
    691
    692
    693
    694
    695
    696
    697
    698
    699
    700
    701
    702
    703
    704
    705
    706
    707
    708
    709
    710
    711
    712
    713
    714
    715
    716
    717
    718
    719
    720
    721
    722
    723
    724
    725
    726
    727
    728
    729
    730
    731
    732
    733
    734
    735
    736
    737
    738
    739
    740
    741
    742
    743
    744
    745
    746
    747
    748
    749
    750
    751
    752
    753
    754
    755
    756
    757
    758
    759
    760
    761
    762
    763
    764
    765
    766
    767
    768
    769
    770
    771
    772
    773
    774
    775
    776
    777
    778
    779
    780
    781
    782
    783
    784
    785
    786
    787
    788
    789
    790
    791
    792
    793
    794
    795
    796
    797
    798
    799
    800
    801
    802
    803
    804
    805
    806
    807
    808
    809
    810
    811
    812
    813
    814
    815
    816
    817
    818
    819
    820
    821
    822
    823
    824
    825
    826
    827
    828
    829
    830
    831
    832
    833
    834
    835
    836
    837
    838
    839
    840
    841
    842
    843
    844
    845
    846
    847
    848
    849
    850
    851
    852
    853
    854
    855
    856
    857
    858
    859
    860
    861
    862
    863
    864
    865
    866
    867
    868
    869
    870
    871
    872
    873
    874
    875
    876
    877
    878
    879
    880
    881
    882
    883
    884
    885
    886
    887
    888
    889
    890
    891
    892
    893
    894
    895
    896
    897
    898
    899
    900
    901
    902
    903
    904
    905
    906
    907
    908
    909
    910
    911
    912
    913
    914
    915
    916
    917
    918
    919
    920
    921
    922
    923
    924
    925
    926
    927
    928
    929
    930
    931
    932
    933
    934
    935
    936
    937
    938
    939
    940
    941
    942
    943
    944
    945
    946
    947
    948
    949
    950
    951
    952
    953
    954
    955
    956
    957
    958
    959
    960
    961
    962
    963
    964
    965
    966
    967
    968
    969
    970
    971
    972
    973
    974
    975
    976
    977
    978
    979
    980
    981
    982
```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: June 18, 2000, 01:53:51 ; Search time 593.22 seconds  
(without alignments)  
12231.247 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_28000\_57000  
Perfect score: 29001  
Sequence: 1 GATAATATTATTAATAT.....ACCTTTGATGAGACATGA 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 433070

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	70	0.2	108	1 X12095	Human biallelic po
C 2	67.2	0.2	100	1 T24892	Human gene signatu
C 3	66	0.2	103	1 T20927	Human gene signatu
C 4	65.2	0.2	108	1 X12095	Human biallelic po
C 5	64	0.2	100	1 X12086	Human biallelic po
C 6	63.6	0.2	100	1 X12087	Human biallelic po
C 7	63.6	0.2	100	1 X12085	Human biallelic po
C 8	62.4	0.2	100	1 T24892	Human gene signatu
C 9	62.6	0.2	103	1 T26213	Human gene signatu
C 10	62.2	0.2	108	1 T26828	Human gene signatu
C 11	61	0.2	99	1 T20931	Human gene signatu
C 12	60.2	0.2	103	1 T20927	Human gene signatu
C 13	59.4	0.2	91	1 T25854	Human gene signatu
C 14	59.4	0.2	108	1 T26828	Human gene signatu
C 15	58.6	0.2	108	1 T25009	Human gene signatu
C 16	57.8	0.2	108	1 T25009	Human gene signatu
C 17	57.2	0.2	91	1 T25854	Human gene signatu
C 18	55.6	0.2	103	1 T26213	Human gene signatu
C 19	55.2	0.2	87	1 T21566	Human gene signatu
C 20	55.2	0.2	93	1 T25688	Human gene signatu
C 21	55.4	0.2	110	1 T25260	Human gene signatu
C 22	55	0.2	93	1 T22572	Human gene signatu
C 23	53.4	0.2	87	1 T21566	Human gene signatu
C 24	53.2	0.2	93	1 T22572	Human gene signatu
C 25	53.4	0.2	97	1 T26728	Human gene signatu
C 26	53	0.2	95	1 T23131	Human gene signatu
C 27	53	0.2	109	1 T23895	Human gene signatu
C 28	52.4	0.2	70	1 N60231	Normal chromosome
C 29	52.6	0.2	93	1 T24259	Human gene signatu
C 30	52.4	0.2	110	1 T25260	Human gene signatu
C 31	52	0.2	69	1 Q29016	Probe to internal
C 32	52	0.2	84	1 T25848	Human gene signatu
C 33	52	0.2	93	1 T25688	Human gene signatu
C 34	52.2	0.2	100	1 X12087	Human biallelic po

35	52.2	0.2	100	1 X12085	Human biallelic po
36	52.2	0.2	100	1 X12086	Human biallelic po
C 37	51.6	0.2	99	1 T23728	Human gene signatu
C 38	51.6	0.2	102	1 T20743	Human gene signatu
C 39	51.4	0.2	110	1 T26288	Human gene signatu
C 40	50.8	0.2	89	1 T23891	Human gene signatu
C 41	50.8	0.2	109	1 T23895	Human gene signatu
C 42	51	0.2	110	1 T26288	Human gene signatu
C 43	49.4	0.2	82	1 T25468	Human gene signatu
C 44	48.8	0.2	69	1 Q29016	Probe to internal
C 45	49	0.2	74	1 T25218	Human gene signatu

ALIGNMENTS

RESULT 1

X12095/c  
ID X12095 standard; DNA; 108 BP.  
AC X12095;  
DT 30-MAR-1999 (first entry)  
DE Human biallelic polymorphic DNA fragment TIGR-A003M18a.  
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;  
KW detection; phenotypic typing; characteristic; infection; hereditary;  
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;  
KW treatment; marker; ss.  
OS Homo sapiens.  
PN W09820165-A2.  
PD 14-MAY-1998.  
PF 05-NOV-1997; U20313.  
PR 06-NOV-1996; US-030455.  
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
PI Hudson T, Lander ES, Wang D;  
DR WPI; 98-286974/25.  
PT New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease  
PS Claim 1: Page 219; 310pp; English.  
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
CC markers which have been isolated using the primers represented in  
CC X09121-X10268. The base occupying the polymorphic site is indicated by  
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tubercous sclerosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.  
SQ Sequence 108 BP; 19 A; 23 C; 28 G; 37 T;

Query Match 0.2%; Score 70; DB 1; Length 108;  
Best Local Similarity 83.3%; Pred. No. 0.0086;  
Matches 90; Conservative 1; Mismatches 16; Indels 1; Gaps 1;

QY	7019	CTGTAATCCAGCAC-TTTGGGAGGCCAAAAGGGCGGATCATTTAGGTTCAGAGTTTCG	7077
Db	108	CTATATCCGACACCTTTGGGAGGCCAAGGACGATCACTTGAAGTCAGAGTTTCG	49
QY	7078	AGACCAAGCTGGCCACATGGTGAACACTCCATCTCTACTAAAAATACA	7125
Db	48	AGACCATCTGGCCACACAYAGAAAAACCTCATCTCTACAAAAAGACA	1

```
RESULT 2
ID T24892/c
AC T24892;
DE 05-NOV-1996 (first entry)
DE Human gene signature HUMG506998.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI: 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1720; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 100 BP; 28 A; 22 C; 25 G; 22 T;

Query Match 0.28; Score 67.2; DB 1; Length 100;
Best Local Similarity 78.8%; Pred. No. 0.023;
Matches 78; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 15731 TTTTGTGTTTGGAGACAGAGTCTCACTATCACCCAGGCTGGAGTGCAGTGGCACAATC 15790
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100 TTTGTTTCTTCAACACAGAGTGTCACTCTCTCACCAGCGGAGTGCAGANGGTGCAATC 41

Qy 15791 TCAGCTCACTGCAACCTGCACCTCCTCGGTTTCAAGGGAT 15829
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40 TCAGCTNATTGCAAAATCTGCTCCAGCGTTTCAAGCGAT 2

RESULT 3
T20927/c
ID T20927;
AC T20927;
DE 24-JUL-1996 (first entry)
DE Human gene signature HUMG502180.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI: 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1720; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 100 BP; 28 A; 22 C; 25 G; 22 T;

Query Match 0.28; Score 67.2; DB 1; Length 100;
Best Local Similarity 78.8%; Pred. No. 0.023;
Matches 78; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 15731 TTTTGTGTTTGGAGACAGAGTCTCACTATCACCCAGGCTGGAGTGCAGTGGCACAATC 15790
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100 TTTGTTTCTTCAACACAGAGTGTCACTCTCTCACCAGCGGAGTGCAGANGGTGCAATC 41

Qy 15791 TCAGCTCACTGCAACCTGCACCTCCTCGGTTTCAAGGGAT 15829
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40 TCAGCTNATTGCAAAATCTGCTCCAGCGTTTCAAGCGAT 2

RESULT 4
X12095
ID X12095 standard; DNA; 108 BP.
AC X12095;
DE 30-MAR-1999 (first entry)
DE Human biallelic polymorphic DNA fragment TIGR-A003M18a.
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; heredity;
KW autoimmune disease; cancer; inflammation; drug; therapy; medication;
KW treatment; marker; ss.
OS Homo sapiens.
PN W09820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; U20313.
PR 06-NOV-1996; US-030455.
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
DR WPI: 98-286974/25.
PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
PS Claim 1; Page 219; 310pp; English.
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC X09121-X10268. The base occupying the polymorphic site is indicated by
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
```







```
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 103 BP; 22 A; 27 C; 21 G; 31 T;

Query Match 0.2%; Score 60.2; DB 1; Length 103;
Best Local Similarity 74.7%; Pred. No. 0.28;
Matches 74; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 21738 APTCCCTGCCTCAGCCTCCCAAGCAGCTGGGATTACAGGTACCTGCCACCATGCTGGT 21797
|| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 2 ATCTCCACCTTCCACCTCCCAAGTAGCTGTGGCTACAGGTGTGTGCCACCATGTCCAGC 61
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 21798 TAATTTTGTATTATTAGTAGAGCGGGGTTTCCACCATG 21836
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 62 TGATTTTGTATTATTAGTAGGACAGTATTCTCCATG 100
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
T25854/c
ID T25854 standard; cDNA to mRNA; 91 BP.
AC T25854;
DF 22-OCT-1996 (first entry)
DE Human gene signature HUMGS08084.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PI (OKUB/) OKUBO K.
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1944; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 91 BP; 18 A; 22 C; 28 G; 18 T;

Query Match 0.2%; Score 59.4; DB 1; Length 91;
Best Local Similarity 77.5%; Pred. No. 0.37;
Matches 69; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 15741 TGAGACAGAGTCTACTCTATACCCAGCTGGAGTGGCGACAAATCTCAGCTCACT 15800
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Dy 90 TGAGACAGAGNTCTACGCTGTCAACNAGCTGGAGCGCGAGGATGCCATCTCAGCTCACT 31
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 15801 GCAACCTGCACCTCTCTGGTTCAGGGAT 15829
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

```
Dy 30 TGAACCCNCTGCCTCCTAGGCTCAAGTGAT 2
RESULT 14
T26828/c
ID T26828 standard; cDNA to mRNA; 108 BP.
AC T26828;
DF 14-NOV-1996 (first entry)
DE Human gene signature HUMGS09078.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PI (OKUB/) OKUBO K.
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 2182; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 108 BP; 18 A; 33 C; 23 G; 28 T;

Query Match 0.2%; Score 59.4; DB 1; Length 108;
Best Local Similarity 83.5%; Pred. No. 0.37;
Matches 66; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 6071 CAGCCGGTGCAGTGGCTCATGACTGTAAATCCAGCACTTTGGGAGGTCGAGGCGGTG 6130
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Dy 79 CAGCCGGGCGTGGTGGCTCATGCTGTAATCCAGCACTATGGGAGCGCGANACGGCGG 20
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 6131 ATCAGGAGTCCAGGAGTTC 6149
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Dy 19 ATGACGAGGTCAGGAGATC 1

RESULT 15
T25009/c
ID T25009 standard; cDNA to mRNA; 108 BP.
AC T25009;
DF 07-NOV-1996 (first entry)
DE Human gene signature HUMGS07131.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PI (OKUB/) OKUBO K.
```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2000, 20:31:38 ; Search time 8513.82 Seconds  
(without alignments)  
13806.673 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_28000\_57000  
Perfect score: 29001  
Sequence: 1 GATATTATTATTAATTAT.....ACCTTTGATGAGAACATGA 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 156056

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	93	0.3	109	30	AA243009	AA243009 zr25h02.s
2	93	0.3	109	84	B17434	B17434 345K2.TVB C
3	92.2	0.3	105	61	A1832832	A1832832 at72909.x
4	91.6	0.3	106	37	AA703692	AA703692 ag81a10.f
5	91.6	0.3	106	105	AQ264176	AQ264176 CITBI-EI-
6	91.4	0.3	109	94	AQ028426	AQ028426 CIT-HSP-2
7	90.2	0.3	103	108	AQ535244	AQ535244 RPCI-11-3
8	89.4	0.3	107	35	AA565533	AA565533 nk42b11.s
9	89	0.3	105	28	AA078003	AA078003 7H12008 C
10	89	0.3	105	61	A1832832	A1832832 at72909.x
11	88.6	0.3	103	84	B48914	B48914 RPCI11-4A12
12	88.6	0.3	103	108	AQ535244	AQ535244 RPCI-11-3
13	88.2	0.3	109	105	AQ265749	AQ265749 CITBI-EI-
14	87.6	0.3	102	94	AQ004934	AQ004934 CIT-HSP-2
15	87.6	0.3	103	38	AA807640	AA807640 nx08b05.s
16	87.6	0.3	110	64	AA083640	AA083640 xc49f02.x
17	87.2	0.3	110	33	AA442529	AA442529 zv68b02.f
18	86.6	0.3	102	30	AA226656	AA226656 nc19f09.s
19	86.8	0.3	106	94	AQ062963	AQ062963 CIT-HSP-2
20	86.8	0.3	110	39	AA897366	AA897366 am06h02.s
21	86.6	0.3	110	94	AQ003188	AQ003188 RPCI11-1D
22	85.8	0.3	105	105	AQ282107	AQ282107 RPCI11-94
23	85.4	0.3	103	94	AQ028649	AQ028649 CIT-HSP-2
24	85.6	0.3	110	33	AA442529	AA442529 zv68b02.f
25	85	0.3	101	39	AA835205	AA835205 ak64h01.s
26	85.2	0.3	106	38	AA812141	AA812141 OB48h02.s
27	85	0.3	110	64	AA083640	AA083640 xc49f02.x
28	84.6	0.3	100	35	AA564832	AA564832 nj22a06.s
29	84.8	0.3	104	105	AQ321855	AQ321855 RPCI11-11
30	84.4	0.3	102	36	AA654562	AA654562 nt75f10.s
31	84.4	0.3	102	84	B48088	B48088 RPCI11-4N6
32	84.6	0.3	103	108	AQ584425	AQ584425 RPCI-11-4
33	84.6	0.3	108	84	B65160	B65160 CIT-HSP-201
34	84.2	0.3	107	24	H67040	H67040 yu68c01.r1
35	84.4	0.3	110	79	AA250394	AA250394 2822460.3
36	84.4	0.3	110	79	AA250394	AA250394 2822460.3
37	83.6	0.3	100	35	AA564832	AA564832 nj22a06.s
38	83.8	0.3	103	108	AQ534922	AQ534922 RPCI-11-3
39	83.8	0.3	105	109	AQ637292	AQ637292 RPCI-11-4
40	83.2	0.3	98	24	H67349	H67349 yu68f10.s1
41	83.4	0.3	103	107	AQ485214	AQ485214 RPCI-11-2
42	83.2	0.3	104	108	AQ544583	AQ544583 CITBI-EI-
43	83.2	0.3	107	33	AA385808	AA385808 EST99495
44	83.2	0.3	107	103	AQ240182	AQ240182 CIT-HSP-2
45	83.4	0.3	109	30	AA243009	AA243009 zr25h02.s

## ALIGNMENTS

```

RESULT 1
AA243009  AA243009  109 bp  mRNA  11-MAR-1998
LOCUS      zr25h02.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens
DEFINITION cDNA clone IMAGE:664467 3', similar to contains Alu repetitive
            element;contains element LTR1 repetitive element ;, mRNA sequence.
ACCESSION  AA243009
VERSION     AA243009.1  GI:1873869
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 109)
AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
            WashU-NCI human EST Project
JOURNAL    Unpublished (1997)
COMMENT    On Dec 3, 1996 this sequence version replaced gi:1126869.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1127 5' end Error: 0.00
Seq primer: -41m13 fwd. 5' from Amersham
High quality sequence stop: 102.
FEATURES
    source
        1..109
        /organism="Homo sapiens"
        /db_xref="GDB:5426481"
        /db_xref="taxon:9606"
        /clone="IMAGE:664467"
        /clone_lib="Stratagene NT2 neuronal precursor 937230"
        /tissue_type="neuroepithelial cells"
        /dev_stage="Ntera-2 neuroepithelial cells"
        /lab_host="SOLR (kanamycin resistant)"
        /note="Organ: brain; Vector: pBluescript SK-; Site:1:
        EcoRI; Site2: XhoI; Cloned unidirectionally. Primer:
        Oligo dt. Uninduced, exponentially growing neuroepithelial
        cells (Ntera-2/ci.D1). Average insert size: 1.0 kb;
        Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG
        3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT  19 a 30 c 30 g 30 t
ORIGIN
Query Match 0.3%; Score 93; DB 30; Length 109;
Best Local Similarity 90.8%; Pred. No. 0.22;
Matches 99; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
Qy 10849 GTATTTTATTAGATGAGGGGTTTCCATGTTAGCCAGGATGGTCTCGATCTCTGACC 10908
Db 1 GTATTTTATTAGATGAGACGGGTTTCCACGCTGTAGCAGGATGGTCTCTGATCTCCCTACC 60
Qy 10909 TCGTGATCCACCGCTTGGCTCCCAAGTCTCGGATTACAGCGCGT 10957
Db 61 TCGTGATCCGCGCCACCTCGGCTCCCAAGTCTCGGATTACAGCGCGT 109

RESULT 2
B17434/c  B17434  109 bp  DNA  GSS  04-JUN-1998
LOCUS      345K2.TVB C1P978SKAI Homo sapiens genomic clone A-345K02, genomic
DEFINITION survey sequence.
ACCESSION  B17434
VERSION     B17434.1  GI:2125183
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 109)
AUTHORS   Adams,M.D., Kelley,J.M., Rounsley,S.R. and Venter,J.C.
            Use of a BAC End Sequence Database for Sequence-Ready Map Building
            Unpublished (1997)
JOURNAL    Other-GSSs: 345K02.TP 345K02.TPB
COMMENT    Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.

```

FEATURES	source	Location/Qualifiers
1. .109		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="A-345K02"
		/clone_lib="CIT978SKA1"
		/sex="Female"
		/cell_type="Fibroblast"
		/note="Vector: pBAC108"; Site_1: HindIII; Site_2: HindIII; Caltech Human BAC Library A1"
BASE COUNT	24 a 30 c 31 q 24 t	

	Query Match	0.38;	Score 93;	DB 84;	Length 109;
	Best Local Similarity	90.88;	Pred. No. 0.22;		
	Matches 99;	Conservative	0;	Mismatches 10;	Indels 0; Gaps 0;
Qy	21813	TAGTAGACAGGGGTTTACCACATGTTGGTCAGGCTGGTCTCGAACTCCTGCACCTCAGGTG	21872		
Db	109	TAGTTGACAGGGGTTTTACCACATGGTGCCAGGCTGGTCTCGAACTCCCGACCTCAGGTG	50		
Qy	21873	ATCTGCCACCTCAGCCTCCCAAGTGTGGGATTACAGGCATGAGCCA	21921		
Db	49	ATCCGCCACATCAGCCTCCCAAGTGTCTAGCATTTAGGTTATGAGCCA	1		

RESULT	3
AI832832	
LOCUS	AI832832 105 bp mRNA EST 13-JUL-1999
DEFINITION	at72g09.x1 Barstead colon HPJRB7 Homo sapiens cDNA clone IMAGE:2377600 3' similar to contains Alu repetitive element; contains element MER22 repetitive element ; mRNA sequence.
ACCESSION	AI832832
VERSION	AI832832.1 GI:5454812
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE  
JOURNAL  
COMMENT

ineisingr, B., winter, J., Wyler, J., Waterston, R., and Wilson, R.  
WashU-NCI Human EST Project  
Unpublished (1997)  
On Dec 20, 1995 this sequence version replaced gi:1133644.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from Gibco.

```

seq primer: 400F from Gluco.
Location/Qualifiers
1. 105
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2377600"
/clone_lib="Barstead colon HPLRB7"
/sex="male"
/dev_stage="adult, age 25"
/lab_host="DH10B (phage resistant)"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site.1: EcoRI; Site.2: NotI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
TGTTACGAATCTCAAGTCGAGCGCGCGCCCTTTTTTTTTTTTTTTTTTTTTT
3'; double-stranded cDNA was ligated to Eco RI adaptors
[5', AATTCTACTAGTAAT 3' and 5' ATTACTAGT 3'], digested
with Not I and cloned into the Not I and Eco RI sites of

```

the modified pT73 vector. Library constructed by Bob Barstead.*									
BASE COUNT	17 a	35 c	27 g	26 t					
ORIGIN									
Query Match	0.3%		Score 92.2;	DB 61;	Length 105;				
Best Local Similarity	92.4%;		Pred. No. 0.28;						
Matches 97;	Conservative	0;	Mismatches	8;	Indels	0;	Gaps	0;	
QY	5364	GAGATGGAGTTTCGCTCTGTTGCCCGAGCTGGAGTGC AATGGCGGGATCTCGGCTCACC	5423						
Db	1	GAGACAGATTTTCGCTCTGTTGCCCGAGCTGGAGTGC AATGGTCGATCTGGCTCACC	60						
QY	5424	GCAACCTCTTACCTCCCGAGTTTCAAGCAATTTCTCCTGCCTCAGCCT	5468						
Db	61	GCAACCTCCACTCCCGGGTTTCAAGCGGATTTCTCCTGCCTCAGCCT	105						

RESULT	4
AA703692	
LOCUS	106 bp mRNA
DEFINITION	aq8la10.r1 Stragatene hNT neuron (#937233) Homo sapiens cDNA clone IMAGE:1140858 5' similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION	AA703692
VERSION	AA703692.1
KEYWORDS	GI:2713610
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 106)
AUTHORS	Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Willson,R. WashU-NCI Human EST Project Unpublished (1997)
TITLE	On Sep 12, 1996 this sequence version replaced gi:1397630.
JOURNAL	
COMMENT	

```

FEATURES
source
nry" quarry sequence stop. 35.
location/Qualifiers
1. .106
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1140858"
/clone_lib="Stratagene hNT neuron (#937233)"
/dev_stage="hNT neurons"
/lab_host="SOLR (kanamycin resistant)"
/note="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dT.
Differentiated, post mitotic hNT neurons. Average insert
size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5',
GAATTCGGCAGCAG 3' -3' adaptor sequence: 5',
CTCAGATTTTTTTTTTTTTTTT 3'
19 a 29 q 29 t
BASE COUNT

```

Query Match 0.3%; Score 91.6; DB 37; Length 106;  
Best Local Similarity 91.5%; Pred. No. 0.32;  
Matches 97; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
Qv 10852 TTTTATTAGATGGGGTTTCACCATCTATGCCAGGATGGTCTCGATCTCTCTGACCTCG 10911

```
Db 1 TTTTGTAGACAGAGGTTTACCGTGTAGCCAGGATGCTCGATCTCTGACCTCG 60
Qy 10912 TGATCCACCGCTTGGCTCCCAAAGTCTGGGATTACAGCGTG 10957
Db 61 TGATCTGCCGCTCAGCTCCCAAAGTCTGGGATTACAGCGTG 106

RESULT 5
LOCUS AQ264176.1 106 bp DNA GSS 27-OCT-1998
DEFINITION CITBI-EI-2509A2.TF CITBI-EI Homo sapiens genomic clone 2509A2,
genomic survey sequence.
ACCESSION AQ264176
VERSION AQ264176.1 GI:3792743
KEYWORDS GSS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 106)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
JOURNAL Unpublished (1998)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..106
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2509A2"
/clone_lib="CITBI-EI"
/sex="male"
/cell_type="sperm"
/notes="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
CalTech Human BAC Library D"
BASE COUNT 25 a 30 c 34 g 17 t
ORIGIN

Query Match 0.3%; Score 91.6; DB 105; Length 106;
Best Local Similarity 91.5%; Pred. No. 0.32;
Matches 97; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 10867 GGGTTTACCATTGATCCAGAGATGCTCGATCCCTGACCTCGATCCACCGCTTT 10926
Db 106 GGGTTTACCATTGATCCAGAGATGCTCGATCCCTGATCCCTGATCCACCGCTC 47

Qy 10927 GGCCTCCCAAGTCTGGGATTACAGGCTGAGCCACCGTGCCCGG 10972
Db 46 GGTCTCCCAAGTCTGGGATTACAGGCTGAGATCTGCGCCCGG 1

RESULT 6
LOCUS AQ028426
DEFINITION CIT-HSP-2313G15.TF CIT-HSP Homo sapiens genomic clone 2313G15,
genomic survey sequence.
ACCESSION AQ028426
```

```
VERSION AQ028426.1 GI:3268648
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 109)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..109
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2313G15"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="sperm"
/notes="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
HindIII"
BASE COUNT 19 a 36 c 25 g 29 t
ORIGIN

Query Match 0.3%; Score 91.4; DB 94; Length 109;
Best Local Similarity 89.9%; Pred. No. 0.33;
Matches 98; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 15733 TTTTGTGAGACAGAGTCTCACTCTATCACCCAGGCTGGAGTGCAGTGCGACAATCTC 15792
Db 1 TTGTTTCTGAGCGGACTCTCACTCTCTCACCCAGGCTGGAGTGCAGTGCGACAATCTG 60

Qy 15793 AGTCACTGCAACCTGCACTCTCTGGGTTCAAGGATTCTCTACCTAA 15841
Db 61 AGTCACTGCAACCTGCACTCTCTGGGTTCAAGGATTCTCTCTGCTCA 109

RESULT 7
LOCUS AQ535244
DEFINITION RPCI-11-317H22.TV RPCI-11 Homo sapiens genomic clone
genomic survey sequence.
ACCESSION AQ535244
VERSION AQ535244.1 GI:4846934
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 103)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
Venter,J.C.
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
```

9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbeetig@org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs ([inforesgen.com](http://inforesgen.com)). BAC end search page: [http://www.tigr.org/tdb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html).  
Seq primer: T7  
Class: BAC ends.

#### FEATURES

source

Location/Qualifiers

1. .103  
/organism="Homo sapiens"  
/db\_xref="GDB:7621533"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-317H22"  
/clone\_lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; RPCI11 Human Male BAC Library"

#### BASE COUNT

ORIGIN

31 a

27 c 27 g 18 t

#### Query Match

Best Local Similarity 92.2%; Score 90.2; DB 108; Length 103;  
Matches 95; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

#### QY

7027 CCAGCATTGGGAGCGCAAAAGGCGATCATTTGAGTCAGGAGTTCGAGACCAAGCC 7086

#### Db

1 CCAGCATTGGGAGCGCAAGAGCGGAGATCATTTGAGTCAGGAGTTCGAGACCAAGCC 60

#### QY

7087 TGCCCAACAFGGTGAACCTCCATCTCTACTACAAAATACAAAA 7129

#### Db

61 TGGCCCAACATGGTGAACCCGCTCTGCTATAAATACAAAA 103

#### RESULT

AA565533/c

LOCUS

AA565533 107 bp mRNA EST 08-SEP-1997

DEFINITION

nk42b11.s1 NCI-CGAP\_GC2 Homo sapiens cDNA clone IMAGE:1016157 3', similar to contains Alu repetitive element,, mRNA sequence.

ACCESSION

AA565533

VERSION

AA565533.1 GI:2337172

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

1 (bases 1 to 107)

TITLE

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

On Sep 12, 1996 this sequence version replaced gi:1393355.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: Stratagene, Inc., David B. Krizman, Ph.D.

CDNA Library Arraying: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

[www-bio.llnl.gov/dbp/image/image.html](http://www-bio.llnl.gov/dbp/image/image.html)

Insert Length: 1661 Std Error: 0.00

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 87.

#### FEATURES

source

Location/Qualifiers

1. .107  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1016157"  
/clone\_lib="NCI-CGAP\_GC2"  
/tissue\_type="germ cell tumor"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dr. Bulk germ cell tumor. 5' adaptor sequence: 5' GAAATCGGCACGAG 3' 3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."

#### BASE COUNT

ORIGIN

22 a 34 c 26 g 25 t

#### Query Match

Best Local Similarity 89.7%; Score 89.4; DB 35; Length 107;  
Matches 96; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

#### QY

6217 TGGTGTGTGCTGTAGTCCAGCTACTCAGAGGCTGGGCGCAGAGAGTAATCGTTGAACCT 6276

#### Db

107 TGGTGTGTGCTGTAAATCCAGCTACTCAGAGGCTGAGGCGAGAGTAATCACTTGAACCT 48

#### QY

6277 GGGAGCGGAGATTGCGATGCGGAGATCGCACCCGACCTCCAG 6323

#### Db

47 GGGAGCGAGAGCTGGCAGTGAGCTGAGATTGAGCCACTGCACCTCCAG 1

#### RESULT

9

AA078003/c

LOCUS

AA078003 105 bp mRNA EST 24-SEP-1999

DEFINITION

7H12D08 Chromosome 7 HeLa cDNA Library Homo sapiens cDNA clone 7H12D08, mRNA sequence.

ACCESSION

AA078003

VERSION

AA078003.1 GI:1837477

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

1 (bases 1 to 105)

TITLE

Touchman, J.W., Bouffard, G.G., Weintraub, L.A., Idol, J.R., Wang, L., Robbins, C.M., Nussbaum, J.C., Lovett, M. and Green, E.D.

JOURNAL

2006 expressed-sequence tags derived from human chromosome 7-enriched cDNA libraries

MEDLINE

Genome Res. 7 (3), 281-292 (1997)

COMMENT

97228905

On Apr 14, 1993 this sequence version replaced gi:693433.

Contact: Eric D. Green

Genome Technology Branch

National Human Genome Research Institute/NIH

49 Convent Dr., MSC4431, Building 49, Room 2A08, Bethesda, MD 20892

Tel: 3014020201

Fax: 3014024735

Email: egreen@nhgri.nih.gov

Plate: 12 row: D column: 08

Seq primer: -21M13 (ABI).

Location/Qualifiers

1. .105

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="7H12D08"

/clone\_lib="Chromosome 7 HeLa cDNA Library"

/sex="female"

/cell\_line="HeLa cell line: ATCC"

/lab\_host="E. coli strain DH5 alpha"

/note="Vector: pAMP10; cDNA was generated from cytoplasmic RNA using a mixture of random DNA hexamers and oligo(dT). From this pool of cDNA, human chromosome 7-enriched cDNA was isolated by direct cDNA selection using chromosome 7



```

RESULT 12
A0535244/c
LOCUS
DEFINITION
  A0535244      103 bp      DNA      GSS      18-MAY-1999
  RPCI-11-317H22.TV RPCI-11 Homo sapiens genomic clone
  RPCI-11-317H22, genomic survey sequence.
ACCESSION
  A0535244
VERSION
  A0535244.1 GI:4846934
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 103)
AUTHORS
  Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
  Venter,J.C.
  Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
  Map Building
  Unpublished (1997)
JOURNAL
  Contact: Shaying Zhao, William Nierman, Mark Adams
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: hbe@tigr.org
  Clones are derived from the human BAC library RPCI-11. For BAC
  library availability, please contact Pieter de Jong
  (pieter@dejong.med.buffalo.edu). Clones may be purchased from
  BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
  Research Genet cs (info@resgen.com). BAC end search page:
  http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
  Seq primer: T7
  Class: BAC ends.
FEATURES
  Location/Qualifiers
  1..103
    /organism="Homo sapiens"
    /db_xref="GDB:7621533"
    /clone_lib="RPCI-11"
    /sex="Male"
    /cell_type="Lymphocytes"
  /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
  RPCI11 Human Male BAC Library"
BASE COUNT      31 a      27 c      27 g      18 t
ORIGIN

Query Match      0.3%; Score 88.6; DB 108; Length 103;
Best Local Similarity 91.3%; Pred. No. 0.69;
Matches 94; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 21801 TTTTGTATTTTATAGACACGGGGTTTCACCATGTTGTGTGAGCTGGTCTGGAACCTCC 21860
|||||
Db 103 TTTTGTATTTATAGCAGACAGCGGGTTTCACCATGTTGTGCCAGCTGGTCTGGAACCTCC 44
|||||

QY 21861 TGACCTCAGGTGATCTGCCACCTCAGCCGCCATGCCAAGCTCTGG 21903
|||||
Db 43 TGACCTCAAGTGATCTGCCCGCTCTTGCCCTCCCAAGTGTCTGG 1
|||||

RESULT 13
A0265749/c
LOCUS
DEFINITION
  A0265749      109 bp      DNA      GSS      27-OCT-1998
  CITBI-EI-2510E2.TR CITBI-EI Homo sapiens genomic clone 2510E2,
  genomic survey sequence.
ACCESSION
  A0265749
VERSION
  A0265749.1 GI:3791503
KEYWORDS
  GSS.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE
  1 (bases 1 to 109)
AUTHORS
  Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
  Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
  Venter,J.C.
  Use of a random human BAC End Sequence Database for Sequence-Ready
  Map Building
  Unpublished (1998)
JOURNAL
  Other-GSSs: CITBI-EI-2510E2.TF
  Contact: Mark Adams
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: mdadams@tigr.org
  Clones are available from Research Genetics (info@resgen.com). BAC
  end search page:
  http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
  Seq primer: M13 Reverse
  Class: BAC ends.
FEATURES
  Location/Qualifiers
  1..109
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="2510E2"
    /clone_lib="CITBI-EI"
    /sex="male"
    /cell_type="sperm"
  /note="Vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;
  CalTech Human BAC Library D"
BASE COUNT      24 a      29 c      26 g      30 t
ORIGIN

Query Match      0.3%; Score 88.2; DB 105; Length 109;
Best Local Similarity 88.1%; Pred. No. 0.74;
Matches 96; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 9841 GCCAACATGGTGAACCTCTCTACTAAATAATACAAAATAGCGGCGCATGGTGGT 9900
|||||
Db 109 GCCAGCATGGTGAACCTCTCTCTACTAGAAATACAAAATAGTCGCGGTGGCG 50
|||||

QY 9901 CAGCGCTGTATCCCGACTACTTGGGAGCTGACACAGGAGAATCGCTT 9949
|||||
Db 49 CATGCGCTGTATCCCGACTACTTGGGAAGCTGAGGCAGGAGCATCACTT 1
|||||

RESULT 14
A0004934
LOCUS
DEFINITION
  A0004934      102 bp      DNA      GSS      27-JUN-1998
  CIT-HSP-2292A10.TF CIT-HSP Homo sapiens genomic clone 2292A10,
  genomic survey sequence.
ACCESSION
  A0004934
VERSION
  A0004934.1 GI:3082379
KEYWORDS
  GSS.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 102)
AUTHORS
  Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
  Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
  Simon,M. and Venter,J.C.
  Use of a random BAC End Sequence Database for Sequence-Ready Map
  Building (1998)
JOURNAL
  Contact: Mark Adams
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: mdadams@tigr.org

```

Clones are available from Research Genetics (info@resgen.com). BAC end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: M13-21;  
Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
1..102  
/organism="Homo sapiens"  
/db\_xref="GDB:7151269"  
/db\_xref="taxon:9606"  
/clone="2292A10"  
/clone\_lib="CIR-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2: HindIII"

BASE COUNT 17 a 39 c 26 g 20 t  
ORIGIN

Query Match 0.3%; Score 87.6; DB 94; Length 102;  
Best Local Similarity 91.2%; Pred. No. 0.89;  
Matches 93; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 5386 GCCCAGGCTGGAGTCAATGGCGGATCTCGGCTCACCGCAACCTCTACCTCCAGGTTTC 5445  
Db 1 GCCCAGGCTGGAGTCTATGGCGGATCTCGGCTCACCGCAACCTCCAGGTTTC 60  
QY 5446 AAGCAATTCCTCTGCTCAGCCTCCCGAGTAGCTGGGATTAC 5487  
Db 61 AAGCGACTCTCTGCTTGGCTTAGGCTCCCGAGTAGCTGGGATTAC 102

RESULT 15  
AA807640

LOCUS AA807640 103 bp mRNA EST 05-MAR-1998  
DEFINITION nx08b05.s1 NCI\_CGAP\_GC3 Homo sapiens cDNA clone IMAGE:1255473 3' similar to contains Alu repetitive element;; mRNA sequence.  
ACCESSION AA807640  
VERSION AA807640.1 GI:2877108  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 103)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2151346.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Insert Length: 774 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 87.

FEATURES  
source

Location/Qualifiers  
1..103  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1255473"  
/clone\_lib="NCI\_CGAP\_GC3"  
/tissue\_type="pooled germ cell tumors"

/lab\_host="DH10B"  
/note="vector: p7T73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from 3 pooled germ cell tumors, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T73 vector. Library is not normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 19 a 27 c 30 g 27 t  
ORIGIN

Query Match 0.3%; Score 87.6; DB 38; Length 103;  
Best Local Similarity 91.2%; Pred. No. 0.88;  
Matches 93; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 10857 ATTAGAGATGGGGTTTCACCATGTTAGCCAGGATGGTCTCGATCTCTCGACCTCGTGATC 10916  
Db 2 AGTAGAGATGGGGTTTCACCGTGTAGCCAGGATGCTCGATCTCTCGACCTCTGTGATC 61  
QY 10917 CACCCGCTTTGGCCCTCCCAAGTCTGGGATTACAGGCGTGA 10958  
Db 62 CGCTCACCTCGGCCCTCCCAAGTCTGGGATTACAGGCGTGA 103

Search completed: June 18, 2000, 04:51:11  
Job time: 317638 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2000, 01:37:37 ; Search time 372.1 Seconds  
(without alignments)  
10130.869 Million cell updates/sec

Title: US-08-852-495c-2\_copy\_28000\_57000

Perfect score: 29001

Sequence: 1 GATAATATTATTAATTAT .....ACCTTTGATGAGAACATGA 29001

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 374504

Minimum DB seq length: 10

Maximum DB seq length: 110

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2.6/ptodata/1/ina/5C\_COMB.seq:\*  
4: /cgn2.6/ptodata/1/ina/5D\_COMB.seq:\*  
5: /cgn2.6/ptodata/1/ina/6\_COMB.seq:\*  
6: /cgn2.6/ptodata/1/ina/PTUS\_COMB.seq:\*  
7: /cgn2.6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	80	0.3	105	4	US-08-481-658B-65
C 2	80	0.3	105	4	US-08-477-504A-65
C 3	80	0.3	105	4	US-08-486-756A-65
C 4	80	0.3	105	4	US-08-485-862B-65
C 5	80	0.3	105	5	US-08-787-739-65
C 6	79.4	0.3	105	4	US-08-481-658B-65
C 7	79.4	0.3	105	4	US-08-477-504A-65
C 8	79.4	0.3	105	4	US-08-486-756A-65
C 9	79.4	0.3	105	4	US-08-485-862B-65
C 10	79.4	0.3	105	5	US-08-787-739-65
C 11	65.4	0.2	84	3	US-08-454-557C-91
C 12	65.4	0.2	84	4	US-08-340-426D-91
C 13	65.4	0.2	84	4	US-08-450-673C-91
C 14	65.4	0.2	84	6	PCT-US95-17111A-91
C 15	62.8	0.2	84	3	US-08-454-557C-91
C 16	62.8	0.2	84	4	US-08-340-426D-91
C 17	62.8	0.2	84	4	US-08-450-673C-91
C 18	62.8	0.2	84	6	PCT-US95-17111A-91
C 19	60.8	0.2	85	3	US-08-454-557C-92
C 20	60.8	0.2	85	4	US-08-340-426D-92
C 21	60.8	0.2	85	4	US-08-450-673C-92
C 22	60.8	0.2	85	6	PCT-US95-17111A-92
C 23	60.4	0.2	78	3	US-08-454-557C-70
C 24	60.4	0.2	78	4	US-08-340-426D-70
C 25	60.4	0.2	78	4	US-08-450-673C-70
C 26	60.4	0.2	78	6	PCT-US95-17111A-70
C 27	58.8	0.2	78	3	US-08-454-557C-70

28	58.8	0.2	78	4	US-08-340-426D-70	Sequence 70, Appl
29	58.8	0.2	78	4	US-08-450-673C-70	Sequence 70, Appl
30	58.8	0.2	78	6	PCT-US95-17111A-70	Sequence 70, Appl
C 31	58	0.2	76	3	US-08-454-557C-69	Sequence 69, Appl
C 32	58	0.2	76	4	US-08-340-426D-69	Sequence 69, Appl
C 33	58	0.2	76	4	US-08-450-673C-69	Sequence 69, Appl
C 34	58	0.2	76	6	PCT-US95-17111A-69	Sequence 69, Appl
C 35	55.2	0.2	60	3	US-08-454-557C-57	Sequence 57, Appl
C 36	55.2	0.2	60	4	US-08-340-426D-57	Sequence 57, Appl
C 37	55.2	0.2	60	4	US-08-450-673C-57	Sequence 57, Appl
C 38	55.2	0.2	60	6	PCT-US95-17111A-57	Sequence 57, Appl
C 39	54.6	0.2	94	5	US-08-750-064-6	Sequence 6, Appl
C 40	53.6	0.2	60	3	US-08-454-557C-60	Sequence 60, Appl
C 41	53.6	0.2	60	4	US-08-340-426D-60	Sequence 60, Appl
C 42	53.6	0.2	60	4	US-08-450-673C-60	Sequence 60, Appl
C 43	53.6	0.2	60	6	PCT-US95-17111A-60	Sequence 60, Appl
C 44	51.6	0.2	76	3	US-08-454-557C-69	Sequence 69, Appl
C 45	51.6	0.2	76	4	US-08-340-426D-69	Sequence 69, Appl

ALIGNMENTS

RESULT 1  
US-08-481-658B-65/c  
; Sequence 65, Application US/08481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,658B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.3E  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-481-658B-65

Query Match 0.3%; Score 80; DB 4; Length 105;  
Best Local Similarity 85.6%; Pred. No. 8.6e-10;

Matches 89; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 6099 ATCCGAGCACTTTGGGAGGTCGAGGCGCTGATCAGGAGGTCAGAGTTCAAGACCAGC 6158  
|||||  
Db 105 ATCCGAGCACTTTGGGAGGCGGAGGCTGGTGATCAGAGGTCAGAGTTTGAGAGCAGC 46  
|||||  
QY 6159 CTGACCAAAATGATGAACCCCTGCTCTACTAAAAATACAACA 6202  
|||||  
Db 45 CTGGCCAATATGTTGAACCCCTGCTCTACTAAAGATGTAAAAA 2

## RESULT 2

US-08-477-504A-65/c  
; Sequence 65, Application US/08477504A  
; Patent No. 5972353

## GENERAL INFORMATION:

; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/477.504A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424

## PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/260,190

## FILING DATE: 15-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.3D  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727

## INFORMATION FOR SEQ ID NO: 65:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-477-504A-65

Query Match 0.3%; Score 80; DB 4; Length 105;

Best Local Similarity 85.6%; Pred. NO. 8.6e-10;

Matches 89; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 6099 ATCCGAGCACTTTGGGAGGTCGAGGCGCTGATCAGGAGGTCAGAGTTCAAGACCAGC 6158  
|||||  
Db 105 ATCCGAGCACTTTGGGAGGCGGAGGCTGGTGATCAGAGGTCAGAGTTTGAGAGCAGC 46  
|||||

QY 6159 CTGACCAAAATGATGAACCCCTGCTCTACTAAAAATACAACA 6202  
|||||

Db 45 CTGGCCAATATGTTGAACCCCTGCTCTACTAAAGATGTAAAAA 2

## RESULT 3

US-08-486-756A-65/c  
; Sequence 65, Application US/08486756A  
; Patent No. 5981711

## GENERAL INFORMATION:

; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/486.756A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424

## PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/260,190

## FILING DATE: 15-JUN-1994

## ATTORNEY/AGENT INFORMATION:

; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.3C  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727

## INFORMATION FOR SEQ ID NO: 65:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-486-756A-65

Query Match 0.3%; Score 80; DB 4; Length 105;

Best Local Similarity 85.6%; Pred. NO. 8.6e-10;

Matches 89; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 6099 ATCCGAGCACTTTGGGAGGTCGAGGCGCTGATCAGGAGGTCAGAGTTCAAGACCAGC 6158  
|||||  
Db 105 ATCCGAGCACTTTGGGAGGCGGAGGCTGGTGATCAGAGGTCAGAGTTTGAGAGCAGC 46  
|||||

QY 6159 CTGACCAAAATGATGAACCCCTGCTCTACTAAAAATACAACA 6202  
|||||

Db 45 CTGGCCAATATGTTGAACCCCTGCTCTACTAAAGATGTAAAAA 2

## RESULT 4

US-08-485-862B-65/c  
; Sequence 65, Application US/08485862B  
; Patent No. 5989638

## GENERAL INFORMATION:

; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court

:/ CITY: Tiburon  
:/ STATE: California  
:/ COUNTRY: USA  
:/ ZIP: 94920  
:/ MEDIUM TYPE: Floppy disk  
:/ COMPUTER: IBM PC compatible  
:/ OPERATING SYSTEM: PC-DOS/MS-DOS  
:/ SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
:/ CURRENT APPLICATION DATA:  
:/ APPLICATION NUMBER: US/08/485,862B  
:/ FILING DATE: 07-JUN-1995  
:/ CLASSIFICATION: 435  
:/ PRIOR APPLICATION DATA:  
:/ APPLICATION NUMBER: US 08/477,504  
:/ FILING DATE: 07-JUN-1995  
:/ APPLICATION NUMBER: US 08/260,190  
:/ FILING DATE: 15-JUN-1994  
:/ ATTORNEY/AGENT INFORMATION:  
:/ NAME: Lauder, Leona L.  
:/ REGISTRATION NUMBER: 30,863  
:/ REFERENCE/DOCKET NUMBER: D-0021.3D  
:/ TELECOMMUNICATION INFORMATION:  
:/ TELEPHONE: 415-435-2034  
:/ TELEFAX: 415-435-0727  
:/ INFORMATION FOR SEQ ID NO: 65:  
:/ SEQUENCE CHARACTERISTICS:  
:/ LENGTH: 105 base pairs  
:/ TYPE: nucleic acid  
:/ STRANDEDNESS: single  
:/ TOPOLOGY: linear  
:/ MOLECULE TYPE: DNA (genomic)  
:/ HYPOTHETICAL: NO  
:/ ANTI-SENSE: NO  
:/ US-08-485-862B-65

Query Match 0.3%; Score 80; DB 4; Length 105;  
Best Local Similarity 85.6%; Pred. No. 8.6e-10;  
Matches 89; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 6099 ATCCAGCAGCTTTGGGAGGTCGAGGAGGCTGATCAGGAGGTCAGGAGTTCAGAGCAGC 6158  
Db 105 ATCCAGCAGCTTTGGGAGGCGGAGGCTGATCAGGAGTTCAGAGCAGC 46

QY 6159 CTCACCAAAATGATGAACCCCTCTCTACTATAAAATACAAACA 6202  
Db 45 CTGGCCAATATGTTGAACCCCTCTCTACTATAAAGATGTAAAAA 2

RESULT 5  
US-08-787-739-65/c  
:/ Sequence 65, Application US/08787739  
:/ Patent No. 6027887  
:/ GENERAL INFORMATION:  
:/ APPLICANT: Zavada, Jan  
:/ APPLICANT: Pastorekova, Silvia  
:/ APPLICANT: Pastorek, Jaromir  
:/ TITLE OF INVENTION: MN Gene and Protein  
:/ NUMBER OF SEQUENCES: 96  
:/ CORRESPONDENCE ADDRESS:  
:/ ADDRESSEE: Leona L. Lauder  
:/ STREET: 369 Pine Street, Suite 610  
:/ CITY: San Francisco  
:/ STATE: California  
:/ COUNTRY: USA  
:/ ZIP: 94104  
:/ MEDIUM TYPE: Floppy disk  
:/ COMPUTER: IBM PC compatible  
:/ OPERATING SYSTEM: PC-DOS/MS-DOS  
:/ SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
:/ CURRENT APPLICATION DATA:

:/ APPLICATION NUMBER: US/08/787,739  
:/ FILING DATE: 24-JAN-1997  
:/ PRIOR APPLICATION DATA:  
:/ APPLICATION NUMBER: US 08/485,049  
:/ FILING DATE: 07-JUN-1995  
:/ PRIOR APPLICATION DATA:  
:/ APPLICATION NUMBER: US 08/486,756  
:/ FILING DATE: 07-JUN-1995  
:/ PRIOR APPLICATION DATA:  
:/ APPLICATION NUMBER: US 08/477,504  
:/ FILING DATE: 07-JUN-1995  
:/ PRIOR APPLICATION DATA:  
:/ APPLICATION NUMBER: US 08/481,658  
:/ FILING DATE: 07-JUN-1995  
:/ PRIOR APPLICATION DATA:  
:/ APPLICATION NUMBER: US 08/485,862  
:/ FILING DATE: 07-JUN-1995  
:/ PRIOR APPLICATION DATA:  
:/ APPLICATION NUMBER: US 08/485,863  
:/ FILING DATE: 07-JUN-1995  
:/ PRIOR APPLICATION DATA:  
:/ APPLICATION NUMBER: US 08/487,077  
:/ FILING DATE: 07-JUN-1995  
:/ ATTORNEY/AGENT INFORMATION:  
:/ NAME: Lauder, Leona L.  
:/ REGISTRATION NUMBER: 30,863  
:/ REFERENCE/DOCKET NUMBER: D-0021.4  
:/ TELECOMMUNICATION INFORMATION:  
:/ TELEPHONE: 415-981-2034  
:/ TELEFAX: 415-981-0332  
:/ INFORMATION FOR SEQ ID NO: 65:  
:/ SEQUENCE CHARACTERISTICS:  
:/ LENGTH: 105 base pairs  
:/ TYPE: nucleic acid  
:/ STRANDEDNESS: double  
:/ TOPOLOGY: linear  
:/ MOLECULE TYPE: DNA (genomic)  
:/ HYPOTHETICAL: NO  
:/ ANTI-SENSE: NO  
:/ US-08-787-739-65

Query Match 0.3%; Score 80; DB 5; Length 105;  
Best Local Similarity 85.6%; Pred. No. 8.6e-10;  
Matches 89; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 6099 ATCCAGCAGCTTTGGGAGGTCGAGGAGGCTGATCAGGAGGTCAGGAGTTCAGAGCAGC 6158  
Db 105 ATCCAGCAGCTTTGGGAGGCGGAGGCTGATCAGGAGTTCAGAGCAGC 46

QY 6159 CTCACCAAAATGATGAACCCCTCTCTACTATAAAATACAAACA 6202  
Db 45 CTGGCCAATATGTTGAACCCCTCTCTACTATAAAGATGTAAAAA 2

RESULT 6  
US-08-481-658B-65  
:/ Sequence 65, Application US/08481658B  
:/ Patent No. 5955075  
:/ GENERAL INFORMATION:  
:/ APPLICANT: Zavada, Jan  
:/ APPLICANT: Pastorekova, Silvia  
:/ APPLICANT: Pastorek, Jaromir  
:/ TITLE OF INVENTION: MN Gene and Protein  
:/ NUMBER OF SEQUENCES: 86  
:/ CORRESPONDENCE ADDRESS:  
:/ ADDRESSEE: Leona L. Lauder  
:/ STREET: 6 Mariposa Court  
:/ CITY: Tiburon  
:/ STATE: California  
:/ COUNTRY: USA  
:/ ZIP: 94920  
:/ COMPUTER READABLE FORM:

```

; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEtical: NO
; ANTI-SENSE: NO
;
US-08-477-504A-65

Query Match 0.3%; Score 79.4; DB 4; Length 105;
Best Local Similarity 84.8%; Pred. No. 1.2e-09;
Matches 89; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 10843 TTTTTCCTATTATTTAGAGTGGGGTTTCACCATCTTAGCCAGGATGGTCTCGATCTC 10902
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1 TTTTTCATCTTTAGTAGAGAGGGTTTCACCATATTGGCCAGGCTGCTCTCAAACTC 60

QY 10903 CTGACCTCGTGATCCACCCCGCTTTGGCCTCCCAAAGTGTCTGGGAT 10947
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 61 CTGACCTTGTGATCCACCGCTCGGGCTCCCAAAGTGTCTGGGAT 105

RESULT 8
US-08-486-756A-65
; Sequence 65, Application US/08486756A
; Patent No. 5981711
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86

```

ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,756A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863

REFERENCE/DOCKET NUMBER: D-0021.3C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727

INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

```
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-486-756A-65

Query Match          0.3%; Score 79.4; DB 4; Length 105;
Best Local Similarity 84.8%; Pred. No. 1.2e-09;
Matches 89; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 10843 TTTTGTATTTTATTAGATGGGTTTCACCATGTTAGCCAGGATGCTCGATCTC 10902
Db      1 TTTTGTATCTTTTAGTAGAGACAGGTTTCACCATATTGGCCAGGCTGCTCTCAAACTC 60

QY 10903 CTGACCTCGTGATCCACCGCTTTGGCCTCCCAAAAGTGTGGGAT 10947
Db      61 CTGACCTTGTGATCCACGACGCTCGGCCTCCCAAAAGTGTGGGAT 105

RESULT 10
US-08-787-739-65
; Sequence 65, Application US/08787739
; Patent No. 6027887
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 96
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 369 Pine Street, Suite 610
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/787,739
; FILING DATE: 24-JAN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,049
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/486,756
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/477,504
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/481,658
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,862
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,863
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,077
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-981-2034
; TELEFAX: 415-981-0332
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-787-739-65

; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-486-756A-65

Query Match          0.3%; Score 79.4; DB 4; Length 105;
Best Local Similarity 84.8%; Pred. No. 1.2e-09;
Matches 89; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 10843 TTTTGTATTTTATTAGATGGGTTTCACCATGTTAGCCAGGATGCTCGATCTC 10902
Db      1 TTTTGTATCTTTTAGTAGAGACAGGTTTCACCATATTGGCCAGGCTGCTCTCAAACTC 60

QY 10903 CTGACCTCGTGATCCACCGCTTTGGCCTCCCAAAAGTGTGGGAT 10947
Db      61 CTGACCTTGTGATCCACGACGCTCGGCCTCCCAAAAGTGTGGGAT 105

RESULT 9
US-08-485-862B-65
; Sequence 65, Application US/08485862B
; Patent No. 5989838
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,862B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/477,504
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-485-862B-65

Query Match          0.3%; Score 79.4; DB 4; Length 105;
Best Local Similarity 84.8%; Pred. No. 1.2e-09;
Matches 89; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
```

Query Match 0.3%; Score 79.4; DB 5; Length 105;  
Best Local Similarity 84.8%; Pred. No. 1.2e-09;  
Matches 89; Conservative 0; Mismatches 16; Indels 0; Gaps 0;  
QY 10843 TTTTGTGATTTTATAGAGATGGGTTTCAACATGTTAGCCAGGATGGTCTCGATCTC 10902  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1 TTTTGTACATCTTAGTAGAGACAGGTTTCAACATATTTGGCAGGCTGCTCAAACTC 60  
QY 10903 CTGACCTGTCATCCACCGCTTTGGCCCTCCAAAGTCTGGGAT 10947  
||||| ||||| ||||| || || ||||| ||||| ||||| ||||| ||||| |||||  
Db 61 CTGACCTGTGTACACCAAGCTCGGCCCTCCAAAGTCTGGGAT 105

RESULT 11  
US-08-454-557C-91  
; Sequence 91, Application US/08454557C  
; Patent No. 5830670  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; TITLE OF INVENTION: of Alzheimer's Disease  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, Steven R.  
; REGISTRATION NUMBER: 36,203  
; REFERENCE/DOCKET NUMBER: 0609.3840003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 91:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 84 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
US-08-454-557C-91

Query Match 0.2%; Score 65.4; DB 3; Length 84;  
Best Local Similarity 86.7%; Pred. No. 2e-06;  
Matches 72; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
QY 10875 CCATGTTAGCCAGGATGTCGATCTCTGACCTCGTGATCCACCGCTTTGGCCTCCC 10934  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1 CCATGTTTCATCAGGCTGGTGTGCGAACTCTGACCTCGTGATCCGCGCCTCAGCCTCCC 60  
QY 10935 AAAGTGCTGGGATTACAGCGGTG 10957  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 61 AAAGTGCTGGGATTACAAAGCGTG 83

RESULT 12  
US-08-340-426D-91  
; Sequence 91, Application US/08340426D  
; Patent No. 5948634

GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; APPLICANT: Wands, Jack R.  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; TITLE OF INVENTION: of Alzheimer's Disease  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/340.426D  
; FILING DATE: 14-NOV-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, Steven R.  
; REGISTRATION NUMBER: 36,203  
; REFERENCE/DOCKET NUMBER: 0609.3840002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 91:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 84 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
US-08-340-426D-91

Query Match 0.2%; Score 65.4; DB 4; Length 84;  
Best Local Similarity 86.7%; Pred. No. 2e-06;  
Matches 72; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
QY 10875 CCATGTTAGCCAGGATGTCGATCTCTGACCTCGTGATCCACCGCTTTGGCCTCCC 10934  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1 CCATGTTTCATCAGGCTGGTGTGCGAACTCTGACCTCGTGATCCGCGCCTCAGCCTCCC 60  
QY 10935 AAAGTGCTGGGATTACAGCGGTG 10957  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 61 AAAGTGCTGGGATTACAAAGCGTG 83

RESULT 13  
US-08-450-673C-91  
; Sequence 91, Application US/08450673C  
; Patent No. 5948888  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; APPLICANT: Wands, Jack R.  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; TITLE OF INVENTION: of Alzheimer's Disease  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25





Search completed: June 18, 2000, 09:57:10  
Job time: 334960 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2000, 09:50:49 ; Search time 17970.9 Seconds  
(without alignments)  
-1569.870 Million cell updates/sec

Title: US-08-852-495c-2\_COPY\_56000\_85000  
Perfect score: 29001  
Sequence: 1 TCCCTTCAGGTCTCCAAGGA.....TGGGGACCAAGTTTAAAG 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 370290

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl:

- 1: gb\_ba1:\*
- 2: gb\_ba2:\*
- 3: gb\_om:\*
- 4: gb\_ov:\*
- 5: gb\_pat:\*
- 6: gb\_ph:\*
- 7: gb\_pl1:\*
- 8: gb\_pl2:\*
- 9: gb\_pr1:\*
- 10: gb\_pr2:\*
- 11: gb\_pr3:\*
- 12: gb\_ro:\*
- 13: gb\_sts:\*
- 14: gb\_sy:\*
- 15: gb\_un:\*
- 16: gb\_vl:\*
- 17: em\_fun:\*
- 18: em\_hum1:\*
- 19: em\_in:\*
- 20: em\_in:\*
- 21: em\_om:\*
- 22: em\_ov:\*
- 23: em\_ov:\*
- 24: em\_pat:\*
- 25: em\_ph:\*
- 26: em\_pl:\*
- 27: em\_ro:\*
- 28: em\_sts:\*
- 29: em\_sy:\*
- 30: em\_un:\*
- 31: em\_vl:\*
- 32: gb\_htg1:\*
- 33: gb\_htg2:\*
- 34: gb\_in1:\*
- 35: gb\_in2:\*
- 36: em\_ba1:\*
- 37: em\_ba2:\*
- 38: em\_hum3:\*
- 39: em\_hum4:\*
- 40: gb\_pr4:\*
- 41: gb\_htg3:\*
- 42: gb\_htg4:\*
- 43: gb\_htg5:\*
- 44: gb\_htg6:\*

- 45: gb\_htg7:\*
- 46: em\_htg1:\*
- 47: em\_htg2:\*
- 48: em\_htg3:\*
- 49: em\_hum5:\*
- 50: gb\_pl3:\*
- 51: gb\_pr5:\*
- 52: gb\_htg8:\*
- 53: gb\_htg9:\*
- 54: gb\_htg10:\*
- 55: gb\_htg11:\*
- 56: gb\_htg12:\*
- 57: gb\_htg13:\*
- 58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	92	0.3	108	10	HSLDLRN2	X05250 Human LDL-r
2	92	0.3	108	10	HSLDLRN2	X05250 Human LDL-r
3	88.6	0.3	107	9	HUMALCE162	M87924 Human carc
4	88	0.3	108	11	HSU67803	U67803 Human small
5	84	0.3	103	9	HUMALCE221	M87896 Human carc
6	83.6	0.3	103	9	HUMALCE221	M87896 Human carc
7	83.8	0.3	104	9	HUMALCE272	M87899 Human carc
8	83	0.3	108	10	HSLDLRD1	X05249 Human LDL-r
9	83	0.3	108	10	HSLDLRD2	X05251 Human LDL-r
10	81.4	0.3	108	10	HSLDLRD1	X05249 Human LDL-r
11	81.4	0.3	108	10	HSLDLRD2	X05251 Human LDL-r
12	78.2	0.3	91	13	HUMUT8164A	L30244 Human STS U
13	78.4	0.3	108	10	HSLDLI12	X05248 Human LDL-r
14	77.8	0.3	108	11	HSU67803	U67803 Human small
15	77.8	0.3	110	9	HUMALCE43	M87900 Human carc
16	76	0.3	104	9	HUMALCE272	M87899 Human carc
17	76.2	0.3	108	11	HSU67808	U67808 Human small
18	76.4	0.3	108	11	HSU67808	U67808 Human small
19	76.4	0.3	110	9	HUMALCE43	M87900 Human carc
20	75.8	0.3	107	9	HUMALCE162	M87924 Human carc
21	75.4	0.3	106	13	G32743	G32743 A009P31 Hum
22	75.2	0.3	108	11	HSU67804	U67804 Human small
23	74.4	0.3	103	13	HS8IC8R	X57789 Human sequ
24	73.4	0.3	103	13	HS8IC8R	X57789 Human sequ
25	72	0.2	90	9	HUMDLRFL	R03555 Human low d
26	72	0.2	108	10	HSLDLI12	X05248 Human LDL-r
27	71.4	0.2	107	11	HSU67806	U67806 Human small
28	71.4	0.2	108	11	HSU67804	U67804 Human small
29	70.8	0.2	91	13	HUMUT8164A	L30244 Human STS U
30	70	0.2	107	11	HSU67806	U67806 Human small
31	70	0.2	108	9	HUMDL03M5	D16965 Human HepG2
32	70.2	0.2	110	11	HSU67807	U67807 Human small
33	70.2	0.2	110	11	HSU67807	U67807 Human small
34	69.4	0.2	97	9	HUMDLR2	M14180 Human low d
35	69.4	0.2	100	9	HUMALNSA	D45223 Human GALNS
36	68.8	0.2	80	9	HUMERKFAE	M36135 Human alpha
37	68.4	0.2	95	13	HUMUT8002B	L30176 Human STS U
38	68.4	0.2	102	13	G32906	G32906 A009W09 Hum
39	67.6	0.2	101	10	S79560	S79560 HRX (intron
40	67.8	0.2	108	9	HUMDL03M5	D16965 Human HepG2
41	67.2	0.2	84	5	AR051521	AR051521 Sequence
42	67.2	0.2	90	9	HUMDLRFL	R03555 Human low d
43	67.4	0.2	102	13	G32906	G32906 A009W09 Hum
44	67	0.2	100	13	HUMUT931A	L31299 Human STS U
45	66	0.2	100	10	HSLAS27	X91545 H.sapiens D

ALIGNMENTS

```

RESULT 1
HSLDLRN2      108 bp      DNA      PRI      20-MAY-1992
LOCUS          Human LDL-receptor gene intron 14 fragment (normal gene).
DEFINITION
ACCESSION      X05250
VERSION        GI:34337
KEYWORDS       Alu repetitive sequence; low density lipoprotein receptor.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
REFERENCE      1 (bases 1 to 108)
AUTHORS        Horsthemke, B., Beisiegel, U., Dunning, A., Havinga, J.R.,
               Williamson, R. and Humphries, S.
TITLE          Unequal crossing-over between two alu-repetitive DNA sequences in
               the low-density-lipoprotein-receptor gene. A possible mechanism for
               the defect in a patient with familial hypercholesterolaemia
JOURNAL        Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE        87161901
COMMENT        See X05252 for deletion junction
               Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES       Location/Qualifiers
               source
               1..108
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               1..108
               /note="intron XIV fragment"
BASE COUNT     28 a 23 c 39 g 18 t
ORIGIN
Query Match    0.3%; Score 92; DB 10; Length 108;
Best Local Similarity 90.7%; Pred. No. 2.2e-06;
Matches 98; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 10935 ACAAAAGTTAGTCGGCGTGGGCACATGCTGTAGTCCAGCTACTGGGAGGCTGAG 10994
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ACAAAATTAGCCAGCGGTGGTGGCAGGTGCTGTAAATCCAGCTACTCGGAGGCTGAG 60

QY 10995 GCAGGAGAATTCCTTGAAGTCCGGAGGCGGAGGTGTCAGTGCAGCCGAG 11042
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GCAGGAGAATTCCTTGAAGTCCGGAGGCGGAGGTGTCAGTGCAGCCGAG 108

RESULT 2
HSLDLRN2/c
LOCUS          Human LDL-receptor gene intron 14 fragment (normal gene).
DEFINITION
ACCESSION      X05250
VERSION        GI:34337
KEYWORDS       Alu repetitive sequence; low density lipoprotein receptor.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
REFERENCE      1 (bases 1 to 108)
AUTHORS        Horsthemke, B., Beisiegel, U., Dunning, A., Havinga, J.R.,
               Williamson, R. and Humphries, S.
TITLE          Unequal crossing-over between two alu-repetitive DNA sequences in
               the low-density-lipoprotein-receptor gene. A possible mechanism for
               the defect in a patient with familial hypercholesterolaemia
JOURNAL        Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE        87161901
COMMENT        See X05252 for deletion junction
               Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES       Location/Qualifiers
               source
               1..108
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               1..108
               /note="intron XIV fragment"
BASE COUNT     28 a 23 c 39 g 18 t
intron
Query Match    0.3%; Score 92; DB 10; Length 108;
Best Local Similarity 90.7%; Pred. No. 2.2e-06;
Matches 98; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 10935 ACAAAAGTTAGTCGGCGTGGGCACATGCTGTAGTCCAGCTACTGGGAGGCTGAG 10994
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ACAAAATTAGCCAGCGGTGGTGGCAGGTGCTGTAAATCCAGCTACTCGGAGGCTGAG 60

QY 10995 GCAGGAGAATTCCTTGAAGTCCGGAGGCGGAGGTGTCAGTGCAGCCGAG 11042
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GCAGGAGAATTCCTTGAAGTCCGGAGGCGGAGGTGTCAGTGCAGCCGAG 108

```

```

ORIGIN
Query Match    0.3%; Score 92; DB 10; Length 108;
Best Local Similarity 90.7%; Pred. No. 2.2e-06;
Matches 98; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 24308 CTCGGCTCACTCAACCTCCGCTCAGGGTTCACAGGATTCTCCTGCTCCGCTCCG 24367
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 108 CTCGGCTCACTCAACCTCTGCTCTCTGGTTCACAGCAATCTCTCTGCTCAGCTCCG 49

QY 24368 AGTAGCTGAGATTACAGGGCTGCCACCATGCCGCTAAATTTTGT 24415
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 48 AGTAGCTGGGATTACAGGCACCTGCCACACGCTGCTAAATTTTGT 1

RESULT 3
HUMALCE162/c
LOCUS          Human carcinoma cell-derived Alu RNA transcript, clone CE162.
DEFINITION
ACCESSION      M87924
VERSION        M87924.1 GI:174871
KEYWORDS       Alu repeat.
SOURCE         Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 107)
AUTHORS        Sinnett, D., Richer, C., Deragon, J.-M. and Labuda, D.
TITLE          Alu RNA transcripts in human embryonal carcinoma cells. Model of
               post-transcriptional selection of master sequences
JOURNAL        J. Mol. Biol. (1992) In press
FEATURES       Location/Qualifiers
               source
               1..107
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /cell_lines="NTERA2D1"
               /dev_stage="embryo"
               /sex="male"
               /tissue_type="carcinoma"
BASE COUNT     28 a 30 c 35 g 14 t
ORIGIN
Query Match    0.3%; Score 88.6; DB 9; Length 107;
Best Local Similarity 91.3%; Pred. No. 8.5e-06;
Matches 94; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 11396 TGTTTGGAGCAGGAGTCTTGTCTGTGCCCAGGCTGAGTGTGTGATCTCCGC 11455
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 107 TTTTGTGAGCAGGAGTCTGCTCTGTGCCCAGGCTGAGTGTGTGATCTCCGC 48

QY 11456 TCACGTGAAGTCCGCTCCCGGATTACGCCATCTCTCTGCC 11498
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 47 TCACGTGAAGTCCGCTCCCGGTTTCAGCCATCTCTCTGCC 5

RESULT 4
HSU67803/c
LOCUS          Human small cytoplasmic Alu transcript.
DEFINITION
ACCESSION      U67803
VERSION        U67803.1 GI:2289917
KEYWORDS       Alu.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 108)
AUTHORS        Shaikh, T.H., Roy, A.M., Kim, J., Batzer, M.A. and Deininger, P.L.
TITLE          cDNAs derived from primary and small cytoplasmic Alu (scAlu)
               transcripts
JOURNAL        J. Mol. Biol. 271 (2), 222-234 (1997)

```



	QY	Db	QY	Db
17750	CCATGTTGGTCAGGCGGGTCTTAAACTCCTGACCTCATGATCGCCCACTCAGCCTCCT	17809		
104	CCATGTTAGCAGGCTGGTCTTGAATCTGTGGCTGCGCAATTCTCCTCCTCAGCCTCCC	45		
17810	AAAGTGCTGGGATTACAGGCGTGAGCCACCCCGCGGCAGCA	17852		
44	AAAGTGCTGGGATTACAGGATTCAGGCACGCGCGCGGCAGCA	2		

RESULT	8
HSIDLRD1	
LOCUS	108 bp DNA
DEFINITION	Human LDL-receptor mutated gene with intron 12 deletion junction.
ACCESSION	X05249
VERSION	X05249.1 GI:34335
KEYWORDS	Alu repetitive sequence; low density lipoprotein receptor.
SOURCE	human.

ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	

REFERENCE	1 (bases 1 to 108)
AUTHORS	Horsthemke, B., Beisiegel, U., Dunning, A., Havinga, J. R.,

**TITLE** Unequal crossing-over between two alu-repetitive DNA sequences in the low-density-lipoprotein-receptor gene. A possible mechanism for the defect in a patient with familial hypercholesterolaemia

**JOURNAL** Eur. J. Biochem. 164 (1), 77-81 (1987)

**WILLIAMSON, R. and HUMPHRIES, S.**

MEDLINE	COMMENT
87161901	<p>*source: hypercholesterol aemia</p> <p>See X05248 for receptor normal gene sequence</p> <p>In the defective LDL receptor gene the deletion occurred between two</p> <p>alu-repetitive sequences, that are in the same direction, the</p> <p>deletion eliminates exons 13 and 14 and changes the reading frame</p> <p>of the resulting spliced mRNA.</p> <p>Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.</p>

FEATURES	Location/Qualifiers
source	1..108
	/organism="Homo sapiens"
	/db_xref="taxon.9606"
	/cell_type="blood leukocytes from a patient with familial"
misc_feature	1..108
	/notes="deletion junction region intron 12/ intron 15"
BASE COUNT	20 a 40 c 20 g 28 t

Query Match	0.3%	Score 83	DB 10	Length 108
Best Local Similarity	86.0%	Pred. No.	7.7e-05	
Matches 92: Conservative	0	Mismatches 15	Indels	0
Matches 92: Gaps	0	Mismatches 15	Indels	0

Qy	24309	TCGGCTCACTGCAACCTCCGGCTCAGGGTTCAAGGATTCTCCTGCCCTCGCCTCCCGA	24368
Db	2	TCGGCTCACCAACACTCTGCCTCTGGGTTCAAAACCAATTTCTGCTCAGCTCAGCCTCCCGA	61

QV 24369 GTAGCTGAGATTACAGGGGGCTGCCACCATGCCCGGCTAATTTTGT 24415

62 GTAGCTGGGATTACAGGCACCTGCCACCACCGCTGGCTAAATTTTGT 108

RESULT	9
HSLDLR2/c	
LOCUS	108 bp DNA
DEFINITION	Human LDL-receptor mutated gene with intron 14 deletion junction.
ACCESSION	X05251
VERSION	X05251.1 GI:34336
KEYWORDS	alu repetitive sequence; low density lipoprotein receptor.
SOURCE	human

ORGANISM	SOURCE
Homo sapiens	human.
Eukaryota. Metazoa. Chordata. Vertebrata. Mammalia. Eutheria.	

REFERENCE  
1 (bases 1 to 108)  
Horsthemke, B., Beisiegel, U., Dunning, A., Havinga, J. R.,  
AUTHORS  
Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 108)  
Horsthemke, B., Beisiegel, U., Dunning, A., Havinga, J. R.,  
AUTHORS  
Primates; Catarrhini; Hominoidea; Homo.

**TITLE** Unequal crossing-over between two alu-repetitive DNA sequences in the low-density-lipoprotein-receptor gene. A possible mechanism for the defect in a patient with familial hypercholesterolaemia

**JOURNAL** Eur. J. Biochem. 164 (1), 77-81 (1987)

**Williamson, R. and Humphries, S.**

SOURCE	Sci: 8; Biochem: 194 (17) 77-81
MEDLINE	87161901
COMMENT	*source: hypercholesterol aemia

See X05250 for corresponding normal gene sequence

**FEATURES**  
Location/Qualifiers

```
location/Qualifiers
1. .108
source
/organism="Homo sapiens"
```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="blood leukocytes from a patient with familial"
1..108
intron

```

BASE COUNT	28 a	20 c	40 g	20 t
/note="intron XIV fragment"				

Query Match	0.3%	Score 83;	DB 10;	Length 108;
Best Local Similarity	86.0%	Pred. No. 7.7e-05;		
Matches 92: Conservative	0:	Mismatches 15;	Indels 0;	Gaps 0;

[illegible][illegible]

QY 24309 GTAGCTGGGATTCACAGGCACCTGCCACCACGCCCTGGCTAATTTTGT 1

RESULT	10
HSL.DI.PD1/C	

LOCUS	108 bp	DNA	PRI
HSLLDRD1			20-MAY-1992
DEFINITION		Human LDL-receptor mutated gene with intron 12 deletion junction.	

DEFINITION	HUMAN LBD	RECEPTOR
ACCESSION	X05249	
VERSION	X05249.1	GI.343335

VERSION  
 X05249.1 GI:34335  
 KEYWORDS  
 Alu repetitive sequence; low density lipoprotein receptor.  
 SOURCE  
 human

SOURCE	ORGANISM	Metazoa:	Chordata:	Mammalia:
human.	<i>Homo sapiens</i>	Enkaryota.	Vertebrata.	Eutheria.

Eukaryota; Metazoa; Chordata; Vertebrata  
 Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 108)  
 REFERENCE

REFERENCE  
AUTHORS  
1 (bases 1 to 108)  
Horsthemke, B., Beisiegel, U., Dunning, A., Havinga, J.R.,  
Williamson P. and Humphries S

**TITLE** Unequal crossing-over between two alu-repetitive DNA sequences in the low-density-lipoprotein-receptor gene. A possible mechanism for the defect in a patient with familial hypercholesterolaemia

**AUTHORS** Williamson, K. and Humphries, S.

**JOURNAL** Eur. J. Biochem. 164 (1), 77-81 (1987)

JOURNAL	Eu: J. Biochem. 104 (1), 77-81 (1987)
MEDLINE	87161901
COMMENT	*source: hypercholesterol aemia

COMMENT  
\*source: hypercholesterolemia  
See X05248 for corresponding normal gene sequence  
In the defective LDL-receptor gene the deletion of

in the defective *mdm-1* receptor gene the deletion occurred between two  $\alpha$ -repetitive sequences, that are in the same direction, the deletion eliminates exons 13 and 14 and changes the reading frame of the resulting spliced mRNA.

**FEATURES**  
Location/Qualifiers

```

FEATURES
location/quadcells
1. .108
source
/organism="Homo sapiens"

```

```
misc feature 1. 108
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="blood leukocytes from a patient with familial"
```

```

, misc_record 1..100
/note="deletion junction re
BASE COUNT 20 a 40 c 20 g 28 t

```

BASE COUNT	ORIGIN
20	2
20	3
20	4
20	5
20	6
20	7
20	8
20	9
20	10
20	11
20	12
20	13
20	14
20	15
20	16
20	17
20	18
20	19
20	20
20	21
20	22
20	23
20	24
20	25
20	26
20	27
20	28
20	29
20	30
20	31
20	32
20	33
20	34
20	35
20	36
20	37
20	38
20	39
20	40
20	41
20	42
20	43
20	44
20	45
20	46
20	47
20	48
20	49
20	50
20	51
20	52
20	53
20	54
20	55
20	56
20	57
20	58
20	59
20	60
20	61
20	62
20	63
20	64
20	65
20	66
20	67
20	68
20	69
20	70
20	71
20	72
20	73
20	74
20	75
20	76
20	77
20	78
20	79
20	80
20	81
20	82
20	83
20	84
20	85
20	86
20	87
20	88
20	89
20	90
20	91
20	92
20	93
20	94
20	95
20	96
20	97
20	98
20	99
20	100



[illegible]



---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2000, 10:13:39 ; Search time 593.49 Seconds  
(without alignments)  
12225.683 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_56000\_85000

Perfect score: 29001

Sequence: 1 TCCCTTCAGGCTCTCAAGGA.....TGGGACCAAAAGTTTTTAAG 29001

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 433070

Minimum DB seq length: 10

Maximum DB seq length: 110

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_36:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	71	0.2	100	T24892	Human gene signatu
C 2	70.4	0.2	100	T24892	Human gene signatu
C 3	66.8	0.2	108	X12095	Human biallelic po
C 4	65.2	0.2	108	T26828	Human gene signatu
C 5	64.8	0.2	108	X12095	Human biallelic po
C 6	64.2	0.2	91	T25854	Human gene signatu
C 7	64.4	0.2	108	T26828	Human gene signatu
C 8	63.6	0.2	103	T26213	Human gene signatu
C 9	62.8	0.2	103	T20927	Human gene signatu
C 10	62.4	0.2	103	T20927	Human gene signatu
C 11	61	0.2	91	T25854	Human gene signatu
C 12	61	0.2	103	T26213	Human gene signatu
C 13	59.6	0.2	100	X12087	Human biallelic po
C 14	59.6	0.2	100	X12085	Human biallelic po
C 15	59.6	0.2	100	X12086	Human biallelic po
C 16	59	0.2	93	T22572	Human gene signatu
C 17	59.2	0.2	108	T25009	Human gene signatu
C 18	58	0.2	110	T26288	Human gene signatu
C 19	57.4	0.2	110	T26288	Human gene signatu
C 20	57	0.2	100	X12087	Human biallelic po
C 21	57	0.2	100	X12085	Human biallelic po
C 22	57	0.2	100	X12086	Human biallelic po
C 23	56	0.2	93	T25854	Human gene signatu
C 24	56.2	0.2	108	T25009	Human gene signatu
C 25	55	0.2	109	T23895	Human gene signatu
C 26	54.4	0.2	109	T23895	Human gene signatu
C 27	53.8	0.2	97	T26728	Human gene signatu
C 28	53.6	0.2	99	T20931	Human gene signatu
C 29	53.2	0.2	93	T22841	Human gene signatu
C 30	53.2	0.2	93	T22572	Human gene signatu
C 31	52.6	0.2	70	N60231	Normal chromosome
C 32	52.2	0.2	97	T26728	Human gene signatu
C 33	51.4	0.2	53	Q33621	Microsatellite seq
C 34	51.6	0.2	75	T22841	Human gene signatu

C 35	51.4	0.2	93	1	T24259	Human gene signatu
C 36	50.8	0.2	87	1	T21566	Human gene signatu
C 37	51	0.2	93	1	T24259	Human gene signatu
C 38	51	0.2	99	1	T20931	Human gene signatu
C 39	50.4	0.2	85	1	T26182	Human gene signatu
C 40	50.4	0.2	106	1	Q95210	Simple tandem repe
C 41	50	0.2	87	1	T21566	Human gene signatu
C 42	49.6	0.2	81	1	T24093	Human gene signatu
C 43	49.6	0.2	84	1	T25848	Human gene signatu
C 44	49.6	0.2	93	1	T25888	Human gene signatu
C 45	49.6	0.2	100	1	T25604	Human gene signatu

ALIGNMENTS

RESULT 1

T24892/c  
ID T24892 standard; cDNA to mRNA; 100 BP.

AC T24892;

DT 05-NOV-1996 (first entry)

DE Human gene signature HUMGS06998.

KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;

KW human; cloning; mapping; non-biased library; diagnosis; detection;

KW cell typing; abnormal cell function; ss.

OS Homo sapiens.

PN W0951472-Al.

PD 01-JUN-1995.

PR 11-NOV-1994; J01916.

PR 12-NOV-1993; JP-355504.

PA (MATS/) MATSUBARA K.

PA (OKUB/) OKUBO K.

PI Matsubara K., Okubo K;

DR WPI; 95-206931/27.

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.

PT for diagnosis of abnormal cell function, by preparing cDNA that

PT reflects relative abundance of corresp. mRNA in specific human

PT tissues

PS Claim 1; Page 1720; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in T19001-T26837 and which is able to hybridise to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost

CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library

CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be

CC determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.

SQ Sequence 100 BP; 28 A; 22 C; 25 G; 22 T;

Query Match

Best Local Similarity 0.2%; Score 71; DB 1; Length 100;

Matches 80; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 17560 TTTTTCCTTCAATAGAGTCTCGCTCTGTCCACCCAGGCTGAGTCGAGTCGGCGCAATCT 17619

Db 99 TTGTTTGTTCACCAAGAGTGTCTACTGTCCACCCAGGCGGAGTGTGCAATCT 40

QY 17620 CAGCTCACTGCAAGTCGCTCTCTGTCCACCCAGGCTGAGTCGAGTCGGCGCAATCT 17657

Db 39 CAGCTNATTCGAATTTTTCCTCCACCCAGGTCACGGAT 2

RESULT 2

T24892

ID T24892 standard; cDNA to mRNA; 100 BP.



```

SQ Sequence 108 BP; 18 A; 33 C; 23 G; 28 T;

Query Match
Best Local Similarity 0.2%; Score 65.2; DB 1; Length 108;
Matches 76; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 24927 ATTAAAAAAGGCTGGCGGTGCTCAGCGCTATATCCAGACTCTTTGG 24986
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 96 AGTTAAGAATAAACACCGCGGTGGTGGCTCATGCTGTAAACCCAGCACTATGG 37

QY 24987 GAGGCTTAGCGGGTGATGACAGAGTCAGGAGTTC 25022
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 36 GAGGCGGANACGGGGATGACGAGGTGAGGAGATC 1

RESULT 5
X12095
ID X12095 standard; DNA; 108 BP.
AC X12095;
DE 30-MAR-1999 (first entry)
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW treatment; marker; ss.
OS Homo sapiens.
PN WQ9820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; 020313.
PR 06-NOV-1996; US-030455.
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
WPI: 98-286974/25.
DR New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
PS Claim 1; Page 219; 310pp; English.
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC X09121-X10268. The base occupying the polymorphic site is indicated by
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC segments can also be used to produce medicaments for the treatment or
CC prophylaxis of such diseases.
SQ Sequence 108 BP; 19 A; 23 C; 28 G; 37 T;

Query Match
Best Local Similarity 0.2%; Score 64.8; DB 1; Length 108;
Matches 86; Conservative 1; Mismatches 18; Indels 1; Gaps 1;

QY 19504 TTTTATAGATAGAGGGTTTTCACATGCTGCCAGGCTGCTCGAACTCCTGACCT 19563
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3 TCTTTTGTAGATAGAGGTTTTCCTRTGTGGCCAGGATGCTCGAACTCCTGACTT 62

QY 19564 CAGGCGATCTGCCCGCTCAGCTCCCAA-GTGCTAGGATTACAG 19608
      || |||| || |||| ||||| ||||| ||||| ||||| ||||| |||||
Db 63 CAAGTGATCGTCTGCTTGGCCCTCCCAAAGTGTCTGGGATTATAG 108
```

```

RESULT 6
T25854
ID T25854 standard; cDNA to mRNA; 91 BP.
AC T25854;
DE 22-OCT-1996 (first entry)
DE Human gene signature HUMGS08084.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
WPI: 95-206931/27.
DR Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1944; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared from
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 91 BP; 18 A; 22 C; 28 G; 18 T;

Query Match
Best Local Similarity 0.2%; Score 64.2; DB 1; Length 91;
Matches 72; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 4066 ATCACTTAACACTAGGAGGAGGCTTGCAGTGCAGTCACACATTCGACTCCAG 4125
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 ATCACTTGACCTAGGAGGAGGCTTCAAGTGAGCTGAGATGGCCTCTCGCTCCAG 61

QY 4126 CCTGGGTGACAGTGTGAGACTCTGTCTCA 4154
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 CCTNGGTGACAGGCTGAGANNCTGTCTCA 90

RESULT 7
T26828
ID T26828 standard; cDNA to mRNA; 108 BP.
AC T26828;
DE 14-NOV-1996 (first entry)
DE Human gene signature HUMGS09078.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
WPI: 95-206931/27.
DR Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT
```

PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues

PS Claim 1: Page 2182; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 108 BP; 18 A; 33 C; 23 G; 28 T;

Query Match 0.2%; Score 64.4; DB 1; Length 108;  
Best Local Similarity 80.4%; Pred. No. 0.12;  
Matches 74; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 11604 GATCTCTGACCTCGTGTGATGCGCGCGCTCAGCCTCCCAAGTGTGGATTACAGAGT 11663  
DB 1 GATCTCTGACCTCGTGTGATGCGCGCGCTCAGCCTCCCAAGTGTGGATTACAGAGT 60

QY 11664 GATCCACTGCGCGCGCGCGCTTTTTTTTTT 11695  
DB 61 GAGCCACCACGCGCGCTGTTTATTCTTAT 92

## RESULT 8

ID T26213 standard; cDNA to mRNA; 103 BP.  
AC T26213;  
DT 13-NOV-1996 (first entry)  
DE Human gene signature HUMGS08452.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN WO9514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI: 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1: Page 2029; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.

SQ Sequence 103 BP; 33 A; 21 C; 25 G; 23 T;

Query Match 0.2%; Score 63.6; DB 1; Length 103;  
Best Local Similarity 76.5%; Pred. No. 0.16;  
Matches 78; Conservative 0; Mismatches 24; Indels 0; Gaps 0;  
QY 10100 GATCTCTGAGCTAGAAAGTTTGGGACGCGAGTGTGATGATGCGACTGCACTCCA 10159  
DB 1 GATCCTTGAGTCCAGGAGTTGGTGTACAGTGAGCTATGATGCGACCACTGCACTCCA 60  
QY 10160 GCCTGGGCAACAATGCCAAATCTCTCTCAAAAACAAAACA 10201  
DB 61 GCCTGGGCAACAATGCCAAATCTCTCTTAAGAAAAAAA 102

## RESULT 9

ID T20927 standard; cDNA to mRNA; 103 BP.  
AC T20927;  
DT 24-JUL-1996 (first entry)  
DE Human gene signature HUMGS02180.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN WO9514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI: 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1: Page 758-759; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 103 BP; 22 A; 27 C; 21 G; 31 T;

Query Match 0.2%; Score 62.8; DB 1; Length 103;  
Best Local Similarity 76.0%; Pred. No. 0.21;  
Matches 76; Conservative 0; Mismatches 24; Indels 0; Gaps 0;  
QY 17655 GATCTCTGCTCAGCCTTCCAGTAGCTGGGATTACAGATGCGACCATGCCGAG 17714  
DB 1 GATCTCTCACCCTCCACCTCCCAAGTAGCTGGGTACAGAGTGTGTGCCACCATGTCCAG 60  
QY 17715 CTAAATTTTGTATTTTATAGACAGACGGAATTTACCATG 17754  
DB 61 CTGATTTTGTATTTTATAGTAGGGACAGTATTTCTCCATG 100

## RESULT 10

ID T20927/c  
ID T20927 standard; cDNA to mRNA; 103 BP.



```
Query Match          0.2%; Score 61; DB 1; Length 103;
Best Local Similarity 75.2%; Pred. No. 0.39;
Matches 76; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 17557 TTCCTTTTTCGAAATGAGATCGCGTCTCACCAGGCTGGAGTGGCGGCAA 17616
      || ||||| || || || || || || || || || || || || || || || ||
Db 102 TTTTTCCTTTTCTTAAAGACATGTTCTTACTCTGTGCGCCAGGCTGGAGTGGCGCA 43

QY 17617 TCTCAGCTACATGCAACGTCGCCCTCCCTGGGTTCAAGTGAT 17657
      || ||||| || || || || || || || || || || || || || || || ||
Db 42 TCATAGCTCACTGTAACACCAAACTCCTGGACTCAAGTGAT 2

RESULT 13
X12087/c
ID X12087 standard; DNA; 100 BP.
AC X12087;
DE Human biallelic polymorphic DNA fragment EST98276a.
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW treatment; marker; ss.
OS Homo sapiens.
PN WO9820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; U20313.
PR (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
PI WPI; 98-286974/25.
PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
PS Claim 1; Page 219; 310pp; English.
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC X09121-X10268. The base occupying the polymorphic site is indicated by
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC segments can also be used to produce medicaments for the treatment or
CC prophylaxis of such diseases.
SQ Sequence 100 BP; 21 A; 25 C; 22 G; 31 T;

Query Match          0.2%; Score 59.6; DB 1; Length 100;
Best Local Similarity 74.0%; Pred. No. 0.63;
Matches 74; Conservative 1; Mismatches 25; Indels 0; Gaps 0;

QY 4856 TGTGGCTCACACCTGTAATCCAGCACCTTTGGAGGCTGAGGCGGCAGATCATCTGAGG 4915
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100 TGTGACTCACACCTATAATCCTGGCACTTTGGAGGCTTAGGAAGGAGGATGTTTGA 41

QY 4916 TCAGAAGTTCCAGACGCGCTGCCCAACATGCGGAACCC 4955
      ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40 CCAGGAGCTCAAGACCATCTCGGAACACATGCAAGACTC 1

RESULT 14
X12085/c
ID X12085 standard; DNA; 100 BP.
AC X12086 standard; DNA; 100 BP.
DE Human biallelic polymorphic DNA fragment EST98276b.
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW treatment; marker; ss.
OS Homo sapiens.
PN WO9820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; U20313.
PR (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
PI WPI; 98-286974/25.
PT New isolated nucleic acid segments from the human genome - used for
```

```
AC X12085;
DE Human biallelic polymorphic DNA fragment EST98276c.
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW treatment; marker; ss.
OS Homo sapiens.
PN WO9820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; U20313.
PR (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
PI WPI; 98-286974/25.
PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
PS Claim 1; Page 218; 310pp; English.
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC X09121-X10268. The base occupying the polymorphic site is indicated by
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC segments can also be used to produce medicaments for the treatment or
CC prophylaxis of such diseases.
SQ Sequence 100 BP; 22 A; 25 C; 22 G; 30 T;

Query Match          0.2%; Score 59.6; DB 1; Length 100;
Best Local Similarity 74.0%; Pred. No. 0.63;
Matches 74; Conservative 1; Mismatches 25; Indels 0; Gaps 0;

QY 4856 TGTGGCTCACACCTGTAATCCAGCACCTTTGGAGGCTGAGGCGGCAGATCATCTGAGG 4915
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100 TGTGACTCACACCTATAATCCTGGCACTTTGGAGGCTTAGGAAGGAGGATGTTTGA 41

QY 4916 TCAGAAGTTCCAGACGCGCTGCCCAACATGCGGAACCC 4955
      ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40 CCAGGAGCTCAAGACCATCTCGGAACACATGCAAGACTC 1

RESULT 15
X12086/c
ID X12086 standard; DNA; 100 BP.
DE Human biallelic polymorphic DNA fragment EST98276b.
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW treatment; marker; ss.
OS Homo sapiens.
PN WO9820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; U20313.
PR (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
PI WPI; 98-286974/25.
PT New isolated nucleic acid segments from the human genome - used for
```



PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease  
PS Claim 1; Page 219; 310pp; English.  
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
CC markers which have been isolated using the primers represented in  
CC X09121-X10268. The base occupying the polymorphic site is indicated by  
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.  
SQ Sequence 100 BP; 21 A; 25 C; 22 G; 31 T;

Query Match 0.2%; Score 59.6; DB 1; Length 100;  
Best Local Similarity 74.0%; Pred. No. 0.63;  
Matches 74; Conservative 1; Mismatches 25; Indels 0; Gaps 0;  
QY 4856 TGTGGTCTACACCTGTATCCAGCACTTTGGAGGCTGAGCGGCGCATCTGTGAGG 4915  
DB 100 TGTGACTCACACCTATATCTTGGCACTTTAGGAGGCTAGGAAGGAGGATTGTTGAAA 41  
QY 4916 TCAGAGTTCCAGACCAGCTGGCCCAACATGGCGCAACCC 4955  
DB 40 CCAGGAGCTCAAGACCATCTCTGGGAACATAGCAAGACTC 1

Search completed: June 18, 2000, 18:17:42  
Job time: 364773 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2000, 04:51:11 ; Search time 8511.85 Seconds  
(without alignments)  
13809.868 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_56000\_85000  
Perfect score: 29001  
Sequence: 1 TCCCTTCAGGCTCTCCAAGGA.....TGGGGACCAAAAGTTTTTAAG 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 156056

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result % Query

## ALIGNMENTS

ORGANISM	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 109)
AUTHORS	Hillier, L., Allen, M., Bowles, L., DeBucque, T., Geisel, G., Jost, S., Krimian, D., Kucaba, T., Lacy, M., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterston, R. and Willson, R.
TITLE	WASHU-NCI human EST project
JOURNAL	Unpublished (1997)
COMMENT	On Dec 3, 1996 this sequence version replaced gi:126869. Contact: Willson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1127 Std Error: 0.00 Seq primer: -4lm13 fwd. ET from Amersham High quality sequence stop: 102.



Db 100 AAGAGCCTGTGGAGCAGCGGCTACTGTGCAACGAAAGCAGCGGCTTCTGGCTCC 41

QY 22066 TTAAACCTCAACAGAGCAGCGCTCCGGGAGCAAGC 22105

Db 40 TTAAACCTCAACAGAGCAGCGCTCCGGGAGCAAGC 1

RESULT 5

LOCUS AA835205 101 bp mRNA EST 23-FEB-1998

DEFINITION ak44h01.s1 Barstead pancreas HPLRB1 Homo sapiens cDNA clone IMAGE:1412689 3' similar to contains Alu repetitive element; contains element KER repetitive element ;, mRNA sequence.

ACCESSION AA835205

VERSION AA835205.1 GI:2908933

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 101)

AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R. WASHU-NCI human EST Project

TITLE Unpublished (1997)

JOURNAL On Nov 29, 1993 this sequence version replaced gi:636191.

COMMENT Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40m13 fwd. ET from Amersham.

FEATURES

source

Location/Qualifiers

1. 101

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1412689"

/clone\_lib="Barstead pancreas HPLRB1"

/sex="female"

/dev\_stage="adult, 34 years"

/lab\_host="DH10B"

/note="Organ: pancreas; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st strand cDNA was primed with a Not I - oligo(df) primer [5', TGTTACGAATCTGAAGTGGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [AATCGAGTCCCTG], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed by Bob Barstead."

BASE COUNT 14 a 36 c 27 g 24 t

ORIGIN

Query Match 0.3%; Score 88.8; DB 39; Length 101;

Best Local Similarity 93.0%; Pred. No. 0.2;

Matches 93; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 11402 GAGACGGAGTCTTGCTCTGTGTCGCCAGCTGAGTGCAGTGTGTGATCTCGGCTCACTG 11461

Db 2 GAGACGGAGTCTCACTCTGTGTCGCCAGCTGAGTGCAGTGTGTGATCTCGGCTCACTG 61

QY 11462 CAAGTCGCCCTCCCGGATTCACGCCATTCCTCCCTCA 11501

Db 62 CAAGTCGCCCTCCCGGATTCACGCCATTCCTCCCTCA 101

RESULT 6

B48914/c

LOCUS B48914 103 bp DNA GSS 08-APR-1999

DEFINITION RPCI11-4A12.TP RPCI-11 Homo sapiens genomic clone RPCI-11-4A12, genomic survey sequence.

ACCESSION B48914

VERSION B48914.1 GI:2601151

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 103)

AUTHORS Adams, M.D., Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.

TITLE Use of BAC End Sequences for Sequence-Ready Map Building

JOURNAL Unpublished (1997)

COMMENT Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: mdamas@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/hungen/bac\_end\_search/bac\_end\_search.html

Seq primer: SP6

Class: BAC ends.

FEATURES

source

Location/Qualifiers

1. 103

/organism="Homo sapiens"

/db\_xref="GDB:7501163"

/db\_xref="taxon:9606"

/clone="RPCI-11-4A12"

/clone\_lib="RPCI-11"

/sex="Male"

/cell\_type="Lymphocytes"

/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI; RPCI11 Human Male BAC Library"

BASE COUNT 30 a 28 c 30 g 15 t

ORIGIN

Query Match 0.3%; Score 88.6; DB 84; Length 103;

Best Local Similarity 91.3%; Pred. No. 0.21;

Matches 94; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 11555 TCTATTTTGTAGATAGCGGGTTTCACTTTGTTAACAGGATGCTCTGATCTCTGAC 11614

Db 103 TGTATTTTGTAGAGAGCGGGTTTCACTTTTACCGGTTTACCGGATGCTCTGATCTCTGAC 44

QY 11615 CTCGTGATCGCGCCGCTCAGCCTCCCAAGTGTGGGATTAC 11657

Db 43 CTCGTGATCGCGCCGCTCAGCCTCCCAAGTGTGGGCTTAC 1

RESULT 7

LOCUS B17434 109 bp DNA GSS 04-JUN-1998

DEFINITION 345K2.TVB CIT978SK1 Homo sapiens genomic clone A-345K02, genomic survey sequence.

ACCESSION B17434

VERSION B17434.1 GI:2125183

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 109)

AUTHORS Adams,M.D., Kelley,J.M., Rounsley,S.R. and Venter,J.C.  
TITLE Use of a BAC End Sequence Database for Sequence-Ready Map Building  
JOURNAL Unpublished (1997)  
COMMENT Other-GSSs: 345K02.TP 345K02.TPB  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadam@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html  
Seq primer: T7  
Class: BAC ends.

FEATURES source  
1. .109  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="A-345K02"  
/clone\_lib="CIT978SKA1"  
/sex="Female"  
/cell\_type="Fibroblast"  
/note="Vector: pBAC108L; Site\_1: HindIII; Site\_2: HindIII;  
CalTech Human BAC Library A1"  
BASE COUNT 24 a 30 c 31 g 24 t  
ORIGIN  
Query Match 0.3%; Score 88.2; DB 84; Length 109;  
Best Local Similarity 88.1%; Pred. No. 0.23;  
Matches 96; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 7539 TGGCTCATGCTGTATCCAGCACCTTTGGGAGGCTGAGGTGATGATCACCTGAGGTT 7598  
||||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1 TGGCTCATGCTGTATCCAGCACCTTTGGGAGGCTGAGGTGATGATCACCTGAGGTC 60  
QY 7599 GGCAGTTTGAGACGACCGCGCAACATGTAACCCCATGCTACTA 7647  
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 61 GGGAGTTCGAGACGACCGCTGGCCACCATGTTGAACCCCGTCTCAACTA 109

RESULT 8  
A0028426  
LOCUS A0028426 109 bp DNA GSS 30-JUN-1998  
DEFINITION CIT-HSP-2313G15.TF CIT-HSP Homo sapiens genomic clone 2313G15,  
genomic survey sequence.  
ACCESSION A0028426  
VERSION A0028426.1 GI:3268648  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 109)  
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,  
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,  
Simon,M. and Venter,J.C.  
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map  
Building (1998)  
JOURNAL Unpublished (1998)  
COMMENT Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadam@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: M13-21

FEATURES source  
Class: BAC ends.  
Location/Qualifiers  
1. .109  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2313G15"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
HindIII"  
BASE COUNT 19 a 36 c 25 g 29 t  
ORIGIN  
Query Match 0.3%; Score 88.2; DB 94; Length 109;  
Best Local Similarity 88.1%; Pred. No. 0.23;  
Matches 96; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 17561 TTTTGTGAAATAGAGTCTGCTGTCCACCCAGGCTGGAGTGCAGTGGCGCAATCTC 17620  
||||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1 TTGTTTTCGAGACGAGCTCTCACTCTGTCTGCCAGCTGGAGTGCAGTGGCGACAGTCTG 60  
QY 17621 AGCTCACTGCAAGCTCCGCTCTCTGGTTCAGTGAATTCCTCCGCTCA 17669  
||||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 61 AGCTCACTGCAACCTCCACCTCTCTGGTTCAGCGATTCCTCTCGCTCA 109

RESULT 9  
A0003188  
LOCUS A0003188 110 bp DNA GSS 14-APR-1999  
DEFINITION RPC111-1D10, TPN RPCI-11 Homo sapiens genomic clone RPCI-11-1D10,  
genomic survey sequence.  
ACCESSION A0003188  
VERSION A0003188.1 GI:3030392  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 110)  
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,  
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and  
Venter,J.C.  
TITLE Use of BAC End Sequences for Sequence-Ready Map Building (1998)  
JOURNAL Unpublished (1998)  
COMMENT Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadam@tigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from  
Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html  
Seq primer: SP6  
Class: BAC ends.  
Location/Qualifiers  
1. .110  
/organism="Homo sapiens"  
/db\_xref="GBD:7500081"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-1D10"  
/clone\_lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPC111 Human Male BAC Library"  
BASE COUNT 22 a 27 c 26 g 35 t

## ORIGIN

Query Match 0.3%; Score 88.2; DB 94; Length 110;  
 Best Local Similarity 88.1%; Pred. No. 0.23; Mismatches 0; Gaps 0;  
 Matches 96; Conservative 0; Indels 0; Gaps 0;

Qy 11551 TTTTGTATTTTGTAGATAGCGGGTTTCACTTTGTTTAAACCAGGATGCTCGCATCTCC 11610  
 Db 2 TTTTGTATTTTGTAGATAGCGGGTTTCACTTTGTTTAAACCAGGATGCTCGCATCTCC 61

Qy 11611 TGACCTCGTGATCGCCCGCTCAGCCCTCCCAAGTGTGGGATTACAG 11659  
 Db 62 TGACCTCGTGATCGCCCGCTCAGCCCTCCCAAGTGTGGGATTACAG 110

## RESULT 10

AQ386882/c 110 bp DNA GSS 21-MAY-1999  
 LOCUS RPI11-13414-TV RPI1-11 Homo sapiens genomic clone RPI1-11-13414,  
 DEFINITION genomic survey sequence.

ACCESSION AQ386882  
 VERSION AQ386882.1 GI:4357905  
 KEYWORDS GSS.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 110)  
 Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and  
 Venter, J.C.

TITLE Use of BAC End Sequences from Library RPI1-11 for Sequence-Ready

## JOURNAL

## COMMENT

Map Building  
 Unpublished (1997)  
 Other\_GSSs: RPI11-13414.TJ  
 Contact: Shaying Zhao, William Nierman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: hbeetigr.org  
 Clones are derived from the human BAC library RPI1-11. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from  
 Research Genetics (info@resgen.com). BAC end search page:  
 http://www.tigr.org/tldb/hungen/bac\_end\_search/bac\_end\_search.html  
 Seq primer: T7  
 Class: BAC ends.

## FEATURES

## source

Location/Qualifiers  
 1..110  
 /organism="Homo sapiens"  
 /db\_xref="GDB:7551267"  
 /db\_xref="taxon:9606"  
 /clone="RPI1-11-13414"  
 /clone\_lib="RPI1-11"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
 RPI11 Human Male BAC Library"  
 BASE COUNT 26 a 26 c 38 g 20 t  
 ORIGIN

Query Match 0.3%; Score 87.6; DB 106; Length 110;  
 Best Local Similarity 87.3%; Pred. No. 0.27;  
 Matches 96; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 19521 GGGTTTCCACGCTGGCCGAGCGTGTCTCGAACCTCGACCTCAGGGGATCTGCCCGCC 19580  
 Db 110 GGGTTTCCACGCTGGCCGAGCGTGTCTCGAACCTCGACCTCAGGGGATCTGCCCGCC 51

Qy 19581 TCAGCCTCCCAAGTGTGATTACAGCGTGTAGCCACCAAGCCTGGCC 19630  
 Db 50 TCAGCCTCCCAAGTGTGATTACAGCGTGTAGCCACCAAGCCTGGCC 1

## RESULT 11

AQ264176/c 106 bp DNA GSS 27-OCT-1998  
 LOCUS CITBI-E1-2509A2.TF CITBI-E1 Homo sapiens genomic clone 2509A2,  
 DEFINITION genomic survey sequence.

ACCESSION AQ264176  
 VERSION AQ264176.1 GI:3792743  
 KEYWORDS GSS.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 106)  
 Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,  
 Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and  
 Venter, J.C.

TITLE Use of a random human BAC End Sequence Database for Sequence-Ready

## JOURNAL

## COMMENT

Map Building  
 Unpublished (1998)  
 Other\_GSSs: CITBI-E1-2509A2.TR  
 Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadams@tigr.org  
 Clones are available from Research Genetics (info@resgen.com). BAC  
 end search page:  
 http://www.tigr.org/tldb/hungen/bac\_end\_search/bac\_end\_search.html.  
 Seq primer: M13-21  
 Class: BAC ends.

## FEATURES

## source

Location/Qualifiers  
 1..106  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="2509A2"  
 /clone\_lib="CITBI-E1"  
 /sex="male"  
 /cell\_type="sperm"  
 /note="Vector: pBelOBAC11; Site\_1: EcoRI; Site\_2: EcoRI;  
 Caltech Human BAC Library D"  
 BASE COUNT 25 a 30 c 34 g 17 t  
 ORIGIN

Query Match 0.3%; Score 86.8; DB 105; Length 106;  
 Best Local Similarity 88.7%; Pred. No. 0.34;  
 Matches 94; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 11574 GGGTTTCACTTTGTTTAAACAGGATGCTCGATCTCTCGACCTCGTGTGATCGCCCGCTC 11633  
 Db 106 GGGTTTCACTTTGTTTAAACAGGATGCTCGATCTCTCGACCTCGTGTGATCGCCCGCTC 47

Qy 11634 AGCCTCCCAAGTGTGGGATTACAGGAGTGAGCCACATGCGCCCGG 11679  
 Db 46 GGTCTCCCAAGTGTGGGATTACAGGAGTGAGCCACATGCGCCCGG 1

## RESULT 12

AA244245 110 bp mRNA EST 20-AUG-1997  
 LOCUS nc07a04.s1 NC1-CCAP\_Prl Homo sapiens cDNA clone IMAGE:1007406  
 DEFINITION similar to contains Alu repetitive element; mRNA sequence.

ACCESSION AA244245  
 VERSION AA244245.1 GI:1875104  
 KEYWORDS EST.  
 SOURCE human.



**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE** 1 (bases 1 to 110)  
**AUTHORS** Chissoe, S., Dietrich, N., DuBuque, T., Favell, A., Gish, W.,  
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
 Marais, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
 Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,  
 and Marra, M.

**TITLE** Generation and analysis of 280,000 human expressed sequence tags  
**JOURNAL** Genome Res. 6 (9), 807-828 (1996)  
**COMMENT** On Jan 24, 1995 this sequence version replaced gi:634306.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuauqui,  
 M.D., Michael Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: David B. Krizman, Ph.D.  
 cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone Distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html

**Seq primer:** -41m13 fwd. ET from Amersham  
**High quality sequence stop:** 90.

**FEATURES** Location/Qualifiers  
 1..110  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1007406"  
 /clone\_lib="NCI-CGAP\_Prl"  
 /sex="Male"  
 /dev\_stage="45 years old"  
 /lab\_host="DH10B"  
 /note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
 strand cDNA was primed with oligo(dT)17 on 50 ng of  
 DNase-treated, total cellular RNA obtained from  
 5,000-10,000 microdissected, histologically normal  
 prostate epithelial cells. Double-stranded cDNA was  
 ligated to EcoRI adaptors, 5 cycles of PCR applied to the  
 cDNA with an adaptor-specific primer, and the resulting  
 PCR product subcloned into pAMP10 by the UDG-cloning  
 method (Life Technologies). Average insert size is 600  
 bp. NOTE: Not directionally cloned. This library was  
 constructed by David Krizman."

**BASE COUNT** 17 a 26 c 28 g 38 t 1 others  
**ORIGIN**

Query Match 0.3%; Score 86.6; DB 30; Length 110;  
 Best Local Similarity 86.4%; Pred. No. 0.35;  
 Matches 95; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 17560 TTTTGTGTAATGAGTCTCGCTCTGTCACCCAGGCTGGAGTCAGTGGCGCAATCT 17619  
 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 1 TTTTGTGAGATGAGTCTTGATCTGTGTCACCCAGGCTGGAGTCAGTGGCGCAATCT 60  
 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 17620 CAGCTCAGTCACAGTCGCGCTCTCGGTTCAAGTGATCTCTCGCTCA 17669  
 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 61 TGCTCAGTCACACCTCTGCTCTCTGGTTCAAGATCTCTCTGCTCA 110  
 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

**RESULT 13**  
**LOCUS** H67040 107 bp mRNA EST 27-OCT-1995  
**DEFINITION** yu68c01.r1 Weizmann Olfactory Epithelium Homo sapiens cDNA clone  
 IMAGE:238944 5' similar to contains Alu repetitive element;; mRNA  
 sequence.

**ACCESSION** H67040  
**VERSION** H67040.1 GI:1025780  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE** 1 (bases 1 to 107)  
**AUTHORS** Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
 Chissoe, S., Dietrich, N., DuBuque, T., Favell, A., Gish, W.,  
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
 Marais, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
 Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,  
 and Marra, M.

**TITLE** Generation and analysis of 280,000 human expressed sequence tags  
**JOURNAL** Genome Res. 6 (9), 807-828 (1996)  
**COMMENT** On Nov 29, 1993 this sequence version replaced gi:429999.  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 High quality sequence stops: 101  
 Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: M13RPI  
 High quality sequence stop: 101.

**FEATURES** Location/Qualifiers  
 1..107  
 /organism="Homo sapiens"  
 /db\_xref="GDB:3864328"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:238944"  
 /clone\_lib="Weizmann Olfactory Epithelium"  
 /sex="Female"  
 /tissue\_type="olfactory epithelium"  
 /dev\_stage="35 year old"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /note="Organ: nose; Vector: pBluescript SK-; Site\_1:  
 EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
 Oligo dT. Olfactory epithelium, normal. Average insert  
 size: 0.8 kb; Uni-ZAP XR Vector. Library constructed by N.  
 Walker, D. Lancet, Weizmann Institute of Science. -5'  
 adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor  
 sequence: 5' CTCGATTTTTTTTTTTT 3' -"

**BASE COUNT** 24 a 37 c 20 g 24 t 2 others  
**ORIGIN**

Query Match 0.3%; Score 85.8; DB 24; Length 107;  
 Best Local Similarity 86.9%; Pred. No. 0.44;  
 Matches 93; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 10943 TAGCTGGCGTGTGGCACATGCTGTAGTCCAGCTACTGGGAGGCTGAGCAGGAGA 11002  
 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 107 TAGCTGGGTGTGTAGCACATGCTGTATTCNAGCTACTCAGNAGGCTGAGGTAGGAGA 48  
 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 11003 ATGCTTGAACCTCGGAGCGGAGGTTGCAGTCAGCGAGATTCGCG 11049  
 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 47 ATCGCTGACCCAGGAGGAGTTGCAGTCAGTCGAGATTGTC 1

**RESULT 14**  
**LOCUS** B65160 108 bp DNA GSS 21-JUN-1998  
**DEFINITION** CIT-HSP-2017G2.TRB CIT-HSP Homo sapiens genomic clone 2017G2,  
 genomic survey sequence.

**ACCESSION** B65160  
**VERSION** B65160.1 GI:2639138  
**KEYWORDS** GSS.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 108)  
**AUTHORS** Adams, M.D., Rounsley, S.D., Field, C.E., Bass, S., Linher, K.,



---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2000, 09:57:10 ; Search time 372.13 Seconds  
(without alignments)  
10130.052 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_56000\_85000  
Perfect score: 29001  
Sequence: 1 TCCCTTCAGGTCCTCCAAGGA.....TGGGACCAAGATTTTAAAG 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 374504

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83.2	0.3	105	4	US-08-481-658B-65
2	83.2	0.3	105	4	US-08-477-504A-65
3	83.2	0.3	105	4	US-08-486-756A-65
4	83.2	0.3	105	4	US-08-485-862B-65
5	83.2	0.3	105	5	US-08-787-739-65
6	81.6	0.3	105	4	US-08-481-658B-65
7	81.6	0.3	105	4	US-08-477-504A-65
8	81.6	0.3	105	4	US-08-486-756A-65
9	81.6	0.3	105	4	US-08-485-862B-65
10	81.6	0.3	105	5	US-08-787-739-65
11	67.2	0.2	84	3	US-08-454-557C-91
12	67.2	0.2	84	4	US-08-340-426D-91
13	67.2	0.2	84	4	US-08-450-673C-91
14	67.2	0.2	84	6	PCT-US95-17111A-91
15	62.2	0.2	84	3	US-08-454-557C-91
16	62.2	0.2	84	4	US-08-340-426D-91
17	62.2	0.2	84	4	US-08-450-673C-91
18	62.2	0.2	84	6	PCT-US95-17111A-91
19	58.8	0.2	78	3	US-08-454-557C-70
20	58.8	0.2	78	4	US-08-340-426D-70
21	58.8	0.2	78	4	US-08-450-673C-70
22	58.8	0.2	78	6	PCT-US95-17111A-70
23	54.6	0.2	85	3	US-08-454-557C-92
24	54.4	0.2	85	3	US-08-454-557C-92
25	54.6	0.2	85	4	US-08-340-426D-92
26	54.4	0.2	85	4	US-08-340-426D-92
27	54.6	0.2	85	4	US-08-450-673C-92

Sequence 92, Appl  
Sequence 92, Appl  
Sequence 92, Appl  
Sequence 60, Appl  
Sequence 60, Appl  
Sequence 60, Appl  
Sequence 60, Appl  
Sequence 69, Appl  
Sequence 69, Appl  
Sequence 69, Appl  
Sequence 69, Appl  
Sequence 69, Appl  
Sequence 66, Appl  
Sequence 66, Appl  
Sequence 66, Appl  
Sequence 66, Appl  
Sequence 66, Appl  
Sequence 70, Appl  
Sequence 70, Appl

ALIGNMENTS

RESULT 1  
US-08-481-658B-65  
; Sequence 65, Application US/08481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,658B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.3E  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-0734  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-481-658B-65

Query Match 0.3%; Score 83.2; DB 4; Length 105;  
Best Local Similarity 87.5%; Pred. No. 1.4e-09;

Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 27950 TTTTGTATTTTATTAGACAGGGTTTCACTATGTTGGCAGGCTGATCTCAAACTCC 28009

Db 2 TTTTACATCTTTAGTAGACAGAGGTTTCACCATTTGGCCAGGCTGCTCTCAAACTCC 61

Qy 28010 TGACCTCATGATCCGCCCTGCGCTTGGCCCTCTCAAAAGTGTCTGGAT 28053

Db 62 TGACCTTGATCCACCAGCCTCGGCCCTCCCAAAAGTGTCTGGAT 105

## RESULT 2

US-08-477-504A-65

; Sequence 65, Application US/08477504A

; Patent No. 5972353

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

; CITY: Tiburon

; STATE: California

; COUNTRY: USA

; ZIP: 94920

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/477,504A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/260,190

; FILING DATE: 15-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lauder, Leona L.

; REGISTRATION NUMBER: 30,863

; REFERENCE/DOCKET NUMBER: D-0021.3D

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-435-2034

; TELEFAX: 415-435-0727

; INFORMATION FOR SEQ ID NO: 65:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 105 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-477-504A-65

Query Match 0.3%; Score 83.2; DB 4; Length 105;

Best Local Similarity 87.5%; Pred. No. 1.4e-09;

Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 27950 TTTTGTATTTTATTAGACAGGGTTTCACTATGTTGGCAGGCTGATCTCAAACTCC 28009

Db 2 TTTTACATCTTTAGTAGACAGAGGTTTCACCATTTGGCCAGGCTGCTCTCAAACTCC 61

Qy 28010 TGACCTCATGATCCGCCCTGCGCTTGGCCCTCTCAAAAGTGTCTGGAT 28053

Db 62 TGACCTTGATCCACCAGCCTCGGCCCTCCCAAAAGTGTCTGGAT 105

## RESULT 3

US-08-486-756A-65

; Sequence 65, Application US/08486756A

; Patent No. 5981711

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

; CITY: Tiburon

; STATE: California

; COUNTRY: USA

; ZIP: 94920

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/486,756A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/260,190

; FILING DATE: 15-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lauder, Leona L.

; REGISTRATION NUMBER: 30,863

; REFERENCE/DOCKET NUMBER: D-0021.3C

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-435-2034

; TELEFAX: 415-435-0727

; INFORMATION FOR SEQ ID NO: 65:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 105 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-486-756A-65

Query Match 0.3%; Score 83.2; DB 4; Length 105;

Best Local Similarity 87.5%; Pred. No. 1.4e-09;

Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 27950 TTTTGTATTTTATTAGACAGGGTTTCACTATGTTGGCAGGCTGATCTCAAACTCC 28009

Db 2 TTTTACATCTTTAGTAGACAGAGGTTTTCACCATTTGGCCAGGCTGCTCTCAAACTCC 61

Qy 28010 TGACCTCATGATCCGCCCTGCGCTTGGCCCTCTCAAAAGTGTCTGGAT 28053

Db 62 TGACCTTGATCCACCAGCCTCGGCCCTCCCAAAAGTGTCTGGAT 105

## RESULT 4

US-08-485-862B-65

; Sequence 65, Application US/08485862B

; Patent No. 5989838

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,862B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-485-862B-65

Query Match 0.3%; Score 83.2; DB 4; Length 105;  
Best Local Similarity 87.5%; Pred. No. 1.4e-09;  
Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 27950 TTTTGTATTTTATTAGACAGGTTTTCACATGTTGGCCAGGCTGATCTCAAACTCC 28009  
Db 2 TTTTACATCTTTAGTAGACAGGTTTTCACATATTTGGCCAGGCTGCTCTCAAACTCC 61  
QY 28010 TGACCTCATGATCCGCTCGCTTGGCTCTCAAAAGTGTGGGAT 28053  
Db 62 TGACCTTGATCCACGCTCGGCTCCCAAAAGTGTGGGAT 105

Query Match 0.3%; Score 83.2; DB 5; Length 105;  
Best Local Similarity 87.5%; Pred. No. 1.4e-09;  
Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 27950 TTTTGTATTTTATTAGACAGGTTTTCACATGTTGGCCAGGCTGATCTCAAACTCC 28009  
Db 2 TTTTACATCTTTAGTAGACAGGTTTTCACATATTTGGCCAGGCTGCTCTCAAACTCC 61  
QY 28010 TGACCTCATGATCCGCTCGCTTGGCTCTCAAAAGTGTGGGAT 28053  
Db 62 TGACCTTGATCCACGCTCGGCTCCCAAAAGTGTGGGAT 105

RESULT 5  
US-08-787-739-65  
; Sequence 65, Application US/08/787739  
; Patent No. 6027887  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 96  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 369 Pine Street, Suite 610  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/787,739  
FILING DATE: 24-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,049  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/486,756  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/481,658  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,862  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,863  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/487,077  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-981-2034  
TELEFAX: 415-981-0332  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-787-739-65

Query Match 0.3%; Score 83.2; DB 5; Length 105;  
Best Local Similarity 87.5%; Pred. No. 1.4e-09;  
Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 27950 TTTTGTATTTTATTAGACAGGTTTTCACATGTTGGCCAGGCTGATCTCAAACTCC 28009  
Db 2 TTTTACATCTTTAGTAGACAGGTTTTCACATATTTGGCCAGGCTGCTCTCAAACTCC 61  
QY 28010 TGACCTCATGATCCGCTCGCTTGGCTCTCAAAAGTGTGGGAT 28053  
Db 62 TGACCTTGATCCACGCTCGGCTCCCAAAAGTGTGGGAT 105

RESULT 6  
US-08-481-658B-65/C  
; Sequence 65, Application US/08/481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/481.658B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-481-658B-65

Query Match 0.3%; Score 81.6; DB 4; Length 105;  
Best Local Similarity 86.5%; Pred. No. 3.1e-09;  
Matches 90; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
Qy 24972 ATCCAGACTCTTTGGAGGCGCTAGCGGTGATCACAGGTCTTCAAGACGAGC 25031  
Db 105 ATCCAGCACTTTGGAGGCGCGAGCTGCTGATCACAAGGTCTTCAAGACGAGC 46  
Qy 25032 CTCGCCAAGATGGTGAATCCGCTCTCTACTATAAAGTATAAAA 25075  
Db 45 CTGGCCAATATGGTGAACCCCTGCTCTACTATAAAGTATAAAA 2

RESULT 7  
US-08-477-504A-65/c  
Sequence 65, Application US/08/477504A  
Patent No. 5972553  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477.504A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:

NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-477-504A-65  
Query Match 0.3%; Score 81.6; DB 4; Length 105;  
Best Local Similarity 86.5%; Pred. No. 3.1e-09;  
Matches 90; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
Qy 24972 ATCCAGACTCTTTGGAGGCGCTAGCGGTGATCACAGGTCTTCAAGACGAGC 25031  
Db 105 ATCCAGCACTTTGGAGGCGCGAGCTGCTGATCACAAGGTCTTCAAGACGAGC 46  
Qy 25032 CTCGCCAAGATGGTGAATCCGCTCTCTACTATAAAGTATAAAA 25075  
Db 45 CTGGCCAATATGGTGAACCCCTGCTCTACTATAAAGTATAAAA 2  
RESULT 8  
US-08-486-756A-65/c  
Sequence 65, Application US/08/486756A  
Patent No. 5981711  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486.756A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear



MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-486-756A-65

Query Match 0.3%; Score 81.6; DB 4; Length 105;  
Best Local Similarity 86.5%; Pred. No. 3.1e-09;  
Matches 90; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 24972 ATCCGAGCTCTTTGGAGGCGCTAGGCGGTGGATCAGAGGTTCAGAGTTCAGAGCAGC 25031  
DB 105 ATCCGAGCAGCTTTGGAGGCGCGAGGCTGGTGGATCACAAGGTTCAGAGTTCAGAGCAGC 46  
QY 25032 CTCGCCAAGATGCTGAATCCCGTCTCTACTATAAAGTATAAAAA 25075  
DB 45 CTGGCCATATGTTGAACCCCTGCTCTACTATAAAGATGTAAAAA 2

RESULT 9  
US-08-485-862B-65/c  
Sequence 65, Application US/08485862B  
Patent No. 5989838  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485.862B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727

INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-485-862B-65

Query Match 0.3%; Score 81.6; DB 4; Length 105;  
Best Local Similarity 86.5%; Pred. No. 3.1e-09;  
Matches 90; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 24972 ATCCGAGCTCTTTGGAGGCGCTAGGCGGTGGATCAGAGGTTCAGAGTTCAGAGCAGC 25031  
DB 105 ATCCGAGCAGCTTTGGAGGCGCGAGGCTGGTGGATCACAAGGTTCAGAGTTCAGAGCAGC 46  
QY 25032 CTCGCCAAGATGCTGAATCCCGTCTCTACTATAAAGTATAAAAA 25075  
DB 45 CTGGCCATATGTTGAACCCCTGCTCTACTATAAAGATGTAAAAA 2

RESULT 10  
US-08-787-739-65/c  
Sequence 65, Application US/08787739  
Patent No. 6027887  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 96  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 369 Pine Street, Suite 610  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/787,739  
FILING DATE: 24-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,049  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/486,756  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/481,658  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,862  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,863  
FILING DATE: 07-JUN-1995

ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.4  
TELEPHONE: 415-981-2034  
TELEFAX: 415-981-0332  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-787-739-65

```
Query Match      0.3%; Score 81.6; DB 5; Length 105;
Best Local Similarity 86.5%; Pred. No. 3.1e-09;
Matches 90; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 24972 ATCCAGCTCTTTGGAGGCTAGCGGTGGATCAGAGTCAAGACGAGC 25031
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105 ATCCAGCAGCTTTGGAGGCGGAGGCTGGTGATCAAGAGTGGAGTTGAGAGCAGC 46
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 25032 CTCGCAAGATGGTGAATCCGCTCTACTATAAAGTATAAAA 25075
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 45 CTGCCAATATGGTGAACCCCTGTCTACTAAAGATGTA AAAA 2

RESULT 11
US-08-454-557C-91
; Sequence 91, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,557C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; US-08-454-557C-91

Query Match      0.2%; Score 67.2; DB 3; Length 84;
Best Local Similarity 90.0%; Pred. No. 4e-06;
Matches 7; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 11595 TGTAAACAGGATGCTCGATCTCTGACCTCGTGATCGCGCGCTCAGCCTCCCAAA 11644
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4 TGTTCATCAGGCTGGTGTGCAACTCTCTGACCTCGTGATCGCGCGCTCAGCCTCCCAAA 63
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 11645 GTGCTGGGATTACAGAGTG 11664
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 64 GTGCTGGGATTACAAAGCGTG 83
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
US-08-340-426D-91
; Sequence 91, Application US/08340426D
; Patent No. 5948634
```

```
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/340,426D
; FILING DATE: 14-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; US-08-340-426D-91

Query Match      0.2%; Score 67.2; DB 4; Length 84;
Best Local Similarity 90.0%; Pred. No. 4e-06;
Matches 72; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 11585 TGTAAACAGGATGCTCGATCTCTGACCTCGTGATCGCGCGCTCAGCCTCCCAAA 11644
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4 TGTTCATCAGGCTGGTGTGCAACTCTCTGACCTCGTGATCGCGCGCTCAGCCTCCCAAA 63
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 11645 GTGCTGGGATTACAGAGTG 11664
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 64 GTGCTGGGATTACAAAGCGTG 83
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
US-08-450-673C-91
; Sequence 91, Application US/08450673C
; Patent No. 5948888
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,673C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
;
US-08-450-673C-91

Query Match 0.2%; Score 67.2; DB 4; Length 84;
Best Local Similarity 90.0%; Pred. No. 4e-06;
Matches 72; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11585 TGTTAACAGGATGGTCTCGATCTCCTGACCTCGTGATCGCGCGCTCAGCCTCCCAAA 11644
Db 4 TGTTCATCAGCGTGGTGTGCAACTCTGACCTCGTGATCGCGCGCTCAGCCTCCCAAA 63

QY 11645 GTCTGGGATTACAGGAGTG 11664
Db 64 GTCTGGGATTACAGCGTG 83

RESULT 14
US-08-450-673C-91
; Sequence 91, Application PC/TUS9517111A
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and
; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
;
US-08-450-673C-91

Query Match 0.2%; Score 62.2; DB 3; Length 84;
Best Local Similarity 84.3%; Pred. No. 5e-05;
Matches 70; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 24962 CAGCGCTATATATCCAGCTCTTTGGGAGGCGCTAGGCGGGTGATCAGAGTCAAGGAGTT 25021
Db 83 CAGCGCTTGTATCCAGCACTTTGGGAGGCTTAGGCGGGTGATCAGAGTCAAGGAGTT 24

QY 25022 CAAGACCAGCCTCGCCCAAGATGG 25044
Db 23 CGACACCAGCCTGATGAACATGG 1
```

```
;
; TOPOLOGY: both
;
PCT-US95-17111A-91

Query Match 0.2%; Score 67.2; DB 6; Length 84;
Best Local Similarity 90.0%; Pred. No. 4e-06;
Matches 72; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11585 TGTTAACAGGATGGTCTCGATCTCCTGACCTCGTGATCGCGCGCTCAGCCTCCCAAA 11644
Db 4 TGTTCATCAGCGTGGTGTGCAACTCTGACCTCGTGATCGCGCGCTCAGCCTCCCAAA 63

QY 11645 GTCTGGGATTACAGGAGTG 11664
Db 64 GTCTGGGATTACAGCGTG 83

RESULT 15
US-08-454-557C-91/c
; Sequence 91, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
;
US-08-454-557C-91

Query Match 0.2%; Score 62.2; DB 3; Length 84;
Best Local Similarity 84.3%; Pred. No. 5e-05;
Matches 70; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 24962 CAGCGCTATATATCCAGCTCTTTGGGAGGCGCTAGGCGGGTGATCAGAGTCAAGGAGTT 25021
Db 83 CAGCGCTTGTATCCAGCACTTTGGGAGGCTTAGGCGGGTGATCAGAGTCAAGGAGTT 24

QY 25022 CAAGACCAGCCTCGCCCAAGATGG 25044
Db 23 CGACACCAGCCTGATGAACATGG 1
```

Search completed: June 18, 2000, 17:46:45  
Job time: 363135 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2000, 17:40:24 ; Search time 17971.1 seconds  
(without alignments)  
-1569.847 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_84000\_113000  
Perfect score: 29001  
Sequence: 1 TCAAACCTCCTGACCTCATGA.....ATAAATATCTTTAAATACC 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 370290

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_p11.\*
- 8: gb\_p12.\*
- 9: gb\_p13.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_vi.\*
- 17: em\_fun.\*
- 18: em\_humi.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_om.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pl.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_vi.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_ba1.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*
- 44: gb\_htg6.\*

- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*
- 50: gb\_pl3.\*
- 51: gb\_pr5.\*
- 52: gb\_htg8.\*
- 53: gb\_htg9.\*
- 54: gb\_htg10.\*
- 55: gb\_htg11.\*
- 56: gb\_htg12.\*
- 57: gb\_htg13.\*
- 58: gb\_htg14.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	96.6	0.3	107	9	HUMALCE162	M87924 Human carc
2	89	0.3	108	11	HSU67803	U67803 Human small
3	87.4	0.3	108	10	HSIDLNR2	X05250 Human LDL-r
4	86.8	0.3	108	10	HSIDLNR2	X05250 Human LDL-r
5	85.4	0.3	107	9	HUMALCE162	M87924 Human carc
6	81.4	0.3	103	9	HUMALCE221	M87896 Human carc
7	81.4	0.3	103	9	HUMALCE221	M87896 Human carc
8	81	0.3	108	11	HSU67803	U67803 Human small
9	80	0.3	108	10	HSIDLNR1	X05249 Human LDL-r
10	80	0.3	108	10	HSIDLNR2	X05251 Human LDL-r
11	79.4	0.3	108	10	HSIDLNR1	X05249 Human LDL-r
12	79.4	0.3	108	10	HSIDLNR2	X05251 Human LDL-r
13	78.2	0.3	108	10	HSIDLNR2	X05248 Human LDL-r
14	77.8	0.3	108	11	HSU67804	U67804 Human small
15	76	0.3	104	9	HUMALCE272	M87899 Human carc
16	75.4	0.3	97	9	HUMDLRA2	M14180 Human low d
17	74.8	0.3	110	9	HUMALCE43	M87900 Human carc
18	73.8	0.3	103	13	HSBIC8R	X57789 Human sequ
19	73	0.3	108	11	HSU67808	U67808 Human small
20	72.6	0.3	106	13	G32743	G32743 A009P31 Hum
21	72.8	0.3	110	11	HSU67807	U67807 Human small
22	72.4	0.2	91	13	HUMUT8164A	L30244 Human STS U
23	72.4	0.2	108	10	HSIDLNR1	X05248 Human LDL-r
24	72	0.2	90	9	HUMDLRFL	K03555 Human low d
25	72.2	0.2	108	9	HUMDLRFL	D16965 Human HepG2
26	71.4	0.2	107	11	HSU67806	U67806 Human small
27	71.4	0.2	108	11	HSU67804	U67804 Human small
28	71.6	0.2	110	11	HSU67807	U67807 Human small
29	71.2	0.2	103	13	HSBIC8R	X57789 Human sequ
30	70.4	0.2	80	9	HUMBRKFAE	M36135 Human alpha
31	69.8	0.2	108	11	HSU67808	U67808 Human small
32	68.4	0.2	95	13	HUMUT8002B	L30176 Human STS U
33	67.6	0.2	84	5	AR051521	AR051521 Sequence
34	67.2	0.2	80	9	HUMBRKFAE	M36135 Human alpha
35	67.4	0.2	94	9	HUMHGAL	M13479 Human alpha
36	67.4	0.2	95	13	HUMUT8002B	L30176 Human STS U
37	66.6	0.2	91	13	HUMUT8164A	L30244 Human STS U
38	66.8	0.2	94	9	HUMHGAL	M13479 Human alpha
39	66	0.2	97	9	HUMDLR1	M14178 Human low d
40	66.2	0.2	108	9	HUMDLR1	D16965 Human HepG2
41	66.2	0.2	108	13	G43535	G43535 WIAT-2393-S
42	65.6	0.2	97	9	HUMDLR1	M14179 Human famil
43	65.2	0.2	79	10	S73203	S73203 ALL-1 (tand
44	64.8	0.2	76	11	AF032287	AF032287 Eulemur m
45	65	0.2	107	11	HSU67806	U67806 Human small

ALIGNMENTS

```

RESULT 1
LOCUS HUMALCE162 107 bp ss-RNA PRI 15-APR-1994
DEFINITION Human carcinoma cell-derived Alu RNA transcript, clone CE162.
ACCESSION M87924
VERSION M87924.1 GI:174871
KEYWORDS Alu repeat.
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 107)
AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
post-transcriptional selection of master sequences
JOURNAL J. Mol. Biol. (1992) In press
FEATURES
    source
        1..107
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /cell_line="Ntera2D1"
            /dev_stage="embryo"
            /sex="male"
            /tissue_type="carcinoma"
BASE COUNT 28 a 30 c 35 g 14 t
ORIGIN

Query Match 0.3%; Score 96.6; DB 9; Length 107;
Best Local Similarity 96.1%; Pred. No. 2.3e-09;
Matches 99; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 9477 GGCAGGAAATGGCTGAACGGGAGCGGAGCTTGAGTGCAGCGGAGATCGCCGACG 9536
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 5 GGCAGGAAATGGCTGAACCGCGGAGCGGAGCTTGAGTGCAGCGGAGATCGCCGAC 64
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 9537 GCACCTCCAGCTGGTGTACAGACGCGAGACTCGCTCTCAAAAAA 9579
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 65 GCACCTCCAGCTGGGCGACAGCGAGACTCGCTCTCAAAAAA 107
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 2
HSU67803
LOCUS HSU67803 108 bp RNA PRI 01-AUG-1997
DEFINITION Human small cytoplasmic Alu transcript.
ACCESSION U67803
VERSION U67803.1 GI:2289917
KEYWORDS Alu.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE cDNAs derived from primary and small cytoplasmic Alu (scAlu)
transcripts
JOURNAL J. Mol. Biol. 271 (2), 222-234 (1997)
MEDLINE 97415756
REFERENCE 2 (bases 1 to 108)
AUTHORS Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The
Children's Hospital of Philadelphia, 1004F Abramson Research
Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA
FEATURES
    source
        1..108
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="TscAlu2"
repeat_region 1..108
            /note="scAlu"
            /rpt_family="Alu"

```

```

BASE COUNT 23 a 39 c 30 g 16 t
ORIGIN

Query Match 0.3%; Score 89; DB 11; Length 108;
Best Local Similarity 94.8%; Pred. No. 7.4e-08;
Matches 92; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 9312 GCCTTAATCCAGCACATTGGGAGCGCGAGCGGGCGGATCACGAGTTCAGGATCGA 9371
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 1 GCCTTAATCCAGCACATTGGGAGCGCGAGCGGGCGGATCACGAGTTCAGGATCGA 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 9372 GACCATCTCTGCTTAACACGATGAAACCCCGTCTCTAC 9408
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 61 GACCATCTCTGCTTAACAGGTGAAACCCCGTCTCTAC 97
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 3
HSLDLRN2/c
LOCUS HSLDLRN2 108 bp DNA PRI 20-MAY-1992
DEFINITION Human LDL-receptor gene intron 14 fragment (normal gene).
ACCESSION X05250
VERSION X05250.1 GI:34337
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
Williamson,R. and Humphries,S.
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT See X05252 for deletion junction
FEATURES
    source
        Location/Qualifiers
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            1..108
                /note="intron XIV fragment"
BASE COUNT 28 a 23 c 39 g 18 t
ORIGIN

Query Match 0.3%; Score 87.4; DB 10; Length 108;
Best Local Similarity 89.5%; Pred. No. 1.5e-07;
Matches 94; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 11837 CTCGGCTCAGCGCAACCTCCGCTCCAGGGTTCAGCAATTCCTCGCTCAGCTCCCC 11896
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 108 CTCGGCTCAGTCAACCTCTGCTCTCGGTTCAAGCAATTCCTCGCTCAGCTCCCG 49
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 11897 AGTAATCGGACTACTGGCAAGCGCCACGCGCTGGCTAATTTT 11941
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 48 AGTAGCTGGGATTACAGGCACCTGCCACCGCTGGCTAATTTT 4
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 4
HSLDLRN2
LOCUS HSLDLRN2 108 bp DNA PRI 20-MAY-1992
DEFINITION Human LDL-receptor gene intron 14 fragment (normal gene).
ACCESSION X05250
VERSION X05250.1 GI:34337
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE 1 (bases 1 to 108)  
AUTHORS Horsthemke, B., Beisiegel, U., Dunning, A., Havinga, J. R.,  
Williamson, R. and Humphries, S.  
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in  
the low-density-lipoprotein-receptor gene. A possible mechanism for  
the defect in a patient with familial hypercholesterolaemia  
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)  
MEDLINE 87161901  
COMMENT See X05252 for deletion junction  
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.  
FEATURES  
source Location/Qualifiers  
1..108  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
intron 1..108  
/note="intron XIV fragment"  
BASE COUNT 28 a 23 c 39 g 18 t  
ORIGIN  
  
Query Match 0.3%; Score 86.8; DB 10; Length 108;  
Best Local Similarity 88.7%; Pred. No. 2e-07; Mismatches 0; Gaps 0;  
Matches 94; Conservative 0; Indels 0; Gaps 0;  
  
QY 2829 AAAAAATTAGCGGGCGTGTGGCGGGCGCTGTAGTCCCGAGCTACTTGGGAGGCTGAGGC 2888  
Db 3 AAAAAATTAGCGGGCGTGTGGCGGGCGAGTGTCTGTAACTCCAGCTACTCGGGAGGCTGAGGC 62  
  
QY 2889 AGGAGAATGGCATGAACCTGGGCGGGAGCTTGCAGTGAGCGCGAG 2934  
Db 63 AGGAGAATGCTTGAACCCAGGAGGAGGTTGCAGTGAGCGCGAG 108  
  
RESULT 5  
HUMALCE162/c  
LOCUS HUMALCE162 107 bp ss-RNA PRI 15-APR-1994  
DEFINITION Human carcinoma cell-derived Alu RNA transcript, clone CE162.  
ACCESSION M87924  
VERSION M87924.1 GI:174871  
KEYWORDS Alu repeat.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 107)  
AUTHORS Sinnett, D., Richer, C., Deragon, J.-M. and Labuda, D.  
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of  
post-transcriptional selection of master sequences  
JOURNAL J. Mol. Biol. (1992) In press  
FEATURES  
source Location/Qualifiers  
1..107  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Ntera2D1"  
/dev\_stage="embryo"  
/sex="male"  
/tissue\_type="carcinoma"  
BASE COUNT 28 a 30 c 35 g 14 t  
ORIGIN  
  
Query Match 0.3%; Score 85.4; DB 9; Length 107;  
Best Local Similarity 89.3%; Pred. No. 3.9e-07; Mismatches 0; Gaps 0;  
Matches 92; Conservative 0; Indels 0; Gaps 0;  
  
QY 10702 TTTTGTGAGACTGAGCTTGTCTGTCAACCCAGGCTGGAGTGAATGGCGGATCTCGGC 10761  
Db 107 TTTTGTGAGAGGAGTCTGCTGTCTGTGCGCCAGGCTGGAGTGGCGGATCTCGGC 48  
  
QY 10762 TCACTGCAACCTCGGTTCCCGAGGTTCAAGCGATTTCTACTGCC 10804  
Db 47 TCACTGCAAGCTCGGCTCCCGGTTCAAGCGATTTCTACTGCC 5

RESULT 6  
HUMALCE221  
LOCUS HUMALCE221 103 bp ss-RNA PRI 15-APR-1994  
DEFINITION Human carcinoma cell-derived Alu RNA transcript, clone CE221.  
ACCESSION M87896  
VERSION M87896.1 GI:174874  
KEYWORDS Alu repeat.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 103)  
AUTHORS Sinnett, D., Richer, C., Deragon, J.-M. and Labuda, D.  
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of  
post-transcriptional selection of master sequences  
JOURNAL J. Mol. Biol. (1992) In press  
FEATURES  
source Location/Qualifiers  
1..103  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Ntera2D1"  
/dev\_stage="embryo"  
/sex="male"  
/tissue\_type="carcinoma"  
BASE COUNT 25 a 27 c 33 g 18 t  
ORIGIN  
  
Query Match 0.3%; Score 81.4; DB 9; Length 103;  
Best Local Similarity 88.9%; Pred. No. 2.4e-06; Mismatches 0; Gaps 0;  
Matches 88; Conservative 0; Indels 0; Gaps 0;  
  
QY 9447 GCTGTAGTCCCGAGCTACTCGGGAGGCTGAGCGAGGAGATGGCGTGAACGGGAGGCG 9506  
Db 5 GCCTGTAATCCCGCTACACGGGAAGCTAAGCGAGGAGATCGTTGAACCCGGGAGGCG 64  
  
QY 9507 GAGCTGCACTGAGCGAGATCGCGCCAGCGCACTCCAG 9545  
Db 65 GAGGTGCACTGAGCGGAGATCGTGCCTTGCACCTCCAG 103  
  
RESULT 7  
HUMALCE221/c  
LOCUS HUMALCE221 103 bp ss-RNA PRI 15-APR-1994  
DEFINITION Human carcinoma cell-derived Alu RNA transcript, clone CE221.  
ACCESSION M87896  
VERSION M87896.1 GI:174874  
KEYWORDS Alu repeat.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 103)  
AUTHORS Sinnett, D., Richer, C., Deragon, J.-M. and Labuda, D.  
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of  
post-transcriptional selection of master sequences  
JOURNAL J. Mol. Biol. (1992) In press  
FEATURES  
source Location/Qualifiers  
1..103  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Ntera2D1"  
/dev\_stage="embryo"  
/sex="male"  
/tissue\_type="carcinoma"  
BASE COUNT 25 a 27 c 33 g 18 t  
ORIGIN  
  
Query Match 0.3%; Score 81.4; DB 9; Length 103;  
Best Local Similarity 88.9%; Pred. No. 2.4e-06; Mismatches 0; Gaps 0;  
Matches 88; Conservative 0; Indels 0; Gaps 0;

QY	16540	CTGAGTGCAGTGCACGATCTCGGCTCACTGCAACCTCTGTCTCCAGGGTTCAAGCGAT	16599
Db	103	CTGGAGTGCAAATGCACGATCTCGGCTCACTGCAACCTCCGGTCCCGGGTTCAAGCGAT	44
QY	16600	TCCTCTCTTCCAGCCTATGGAGTAGTACGTGGGATTACAGGC	16638
Db	43	TCCTCTGCCCTTAGCTTCCCGTGTAGCTGGGATTACAGGC	5

RESULT	8
HSU67803/c	
LOCUS	HSU67803.1 108 bp RNA
DEFINITION	Human small cytoplasmic Alu transcript.
ACCESSION	U67803
VERSION	U67803.1 GI:2289917
KEYWORDS	alu.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 108)
AUTHORS	Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE	cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts
JOURNAL	J. Mol. Biol. 271 (2), 222-234 (1997)
MEDLINE	97415756
REFERENCE	2 (bases 1 to 108)
AUTHORS	Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE	Direct Submission
JOURNAL	Submitted (22-AUG-1996) Human Genetics and Molecular Biology. The

FEATURES  
Location/Qualifiers  
source  
1. .108

```

repeat_region
1. .108
/ clones = scahu2
/ note = "scahu"
/ rpt_family = "Alu"
/ rpt_type = dispersed
BASE COUNT      23 a      39 c      30 g      16 t
ORIGIN

Query Match      0.3%; Score 81; DB 11; Length 108;
Best Local Similarity 89.7%; Pred. No. 2.9e-06;
Matches 87; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```

QY	10871	GTAGAGACAGGGTTTACCGGTGTTGGCCAGAGTCTTCTCAATCTCCCTTACCTCTGTGATCC	10930
Db	97	GTAGAGACGGGGTTTACCTTGTGTAGCCAGAGTGTCTCGATCTCTGACCTCTGTGATCC	38
QY	10931	GCCCGCTCTCTCTGCCAAAGTGTCTCGATTACAGAC	10967
Db	37	GCCCGCTCTCGGCTTCCCAAAGTGTCTGGGATTACAGGC	1

RESULT	9
LOCUS	HSLDLRD1
DEFINITION	Human LDL-receptor mutated gene with intron 12 deletion junction.
ACCESSION	X05249
VERSION	X05249.1 GI:34335
KEYWORDS	Alu repetitive sequence; low density lipoprotein receptor.
SOURCE	human.
ORGANISM	homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS	Primates; Catarrhini; Hominiidae; Homo.
	1 (bases 1 to 108)
	Horsthenke,B., Beislegel,U., Dunning,A., Haviga,J.R.,
	Williamson,R. and Humphries,S.

TITLE	Unequal crossing-over between two alu-repetitive DNA sequences in the low-density-lipoprotein-receptor gene. A possible mechanism for the defect in a patient with familial hypercholesterolaemia
JOURNAL	Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE	87161901
COMMENT	*source: hypercholesterol aemia See X05248 for corresponding normal gene sequence In the defective LDL-receptor gene the deletion occurred between two alu-repetitive sequences, that are in the same direction, the deletion eliminates exons 13 and 14 and changes the reading frame of the resulting spliced mRNA. Data kindly reviewed (07-DEC-1987) by HUMPHRIES S. Location/Qualifiers 1. .108 /organism="Homo sapiens" /db_xref="taxon:9606" /cell_type="blood leukocytes from a patient with familial"
FEATURES	
source	1. .108 /organism="Homo sapiens" /db_xref="taxon:9606" /cell_type="blood leukocytes from a patient with familial"
misc_feature	1. .108 /notes="deletion junction region intron 12/ intron 15"
BASE COUNT	20 a 40 c 20 g 28 t
ORIGIN	

	Query Match	0.3%	Score 80;	DB 10;	Length 108;
	Best Local Similarity	85.6%;	Pred. No. 4.6e-06;		
	Matches 8;	Conservative 0;	Mismatches 15;	Indels 0;	Gaps 0;
Qy	11838	TCGCCTCACCGCAACCTCCGCCTCCACGGTTCAAGCAATTTCTCTCGCTCAGCGCTCCCCA	11897		
Db	2	TCGCCTCACCAACACCTCTCGCTCTGGGTTCAACCAATTTTCTGTGCTCAGCGCTCCCGA	61		
Qy	11898	GTAATTGGGACTACTGGCAAGCGCCACAGCGCTGGCTAATTTT	11941		
Db	62	GTAGCTGGGATTACAGGCACCTTGGCCACAGCGCTGGCTAATTTT	105		

RESULT	10
HSLDRD2/c	
LOCUS	
DEFINITION	HSLDRD2 108 bp DNA PRI 20-MAY-1992
ACCESSION	Human LDL-receptor mutated gene with intron 14 deletion junction.
X05251	
VERSION	X05251.1 GI:34336
KEYWORDS	Alu repetitive sequence; low density lipoprotein receptor.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
	Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 108)
AUTHORS	Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R., Williamson,R. and Humphries,S.

```

TITLE      Unequal crossing-over between two alu-repetitive DNA sequences in
           the low-density lipoprotein-receptor gene. A possible mechanism for
           the defect in a patient with familial hypercholesterolaemia
JOURNAL    Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE    87161901
COMMENT    *source: hypercholesterol aemia
           See X05250 for corresponding normal gene sequence
           In the defective LDL-receptor gene the deletion occurred between two
           alu-repetitive sequences, that are in the same direction, the
           deletion eliminates exons 13 and 14 and changes the reading frame
           of the resulting spliced mRNA.
           Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES   Location/Qualifiers
            source          1..108
                           /organism="Homo sapiens"
                           /db_xref="taxon:9606"
                           /cell_type="Blood leukocytes from a patient with familial"
            intron          1..108
                           /note="intron xiv fragment"
BASE COUNT      28 a    20 c    40 g    20 t
ORIGIN

```



	Query Match	0.3%; Score 80; DB 10; Length 108;
	Best Local Similarity	85.6%; Pred. No. 4.6e-06;
	Matches	89; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 11838	TCCGCTCACCGCAACCTCCGCCTCAGGGTTTAAAGCATTCCTGCCGTGAGCTCCCACA 11897   Db 107 TCCGCTCACCAACTCTCGCTCTTG GTTCAAACATTTCTTCCTGCCGTGAGCTCCCAGA 48 	
QY 11898	GTAATTGGGAATACTGGCAAGCGCCACCACGCGCTGGCTAATTTTT 11941   Db 47 GTAGCTGGATTACAGCACCTGCCACACGCGCTGGCTAATTTTT 4	
RESULT 11	HSLDLRD1 108 bp DNA PRI 20-MAY-1992	
LOCUS	Human LDL-receptor mutated gene with intron 12 deletion junction.	
DEFINITION	X05249	
ACCESSION	X05249.1 GI:34335	
VERSION	Alu repetitive sequence; low density lipoprotein receptor. human.	
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
REFERENCE	1 (bases 1 to 108)	
AUTHORS	Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R., Williamson,R. and Humphries,S.	
TITLE	Unequal crossing-over between two alu-repetitive DNA sequences in the low-density-lipoprotein-receptor gene. A possible mechanism for the defect in a patient with familial hypercholesterolaemia	
JOURNAL	Eur. J. Biochem. 164 (1), 77-81 (1987)	
MEDLINE	87161901	
COMMENT	*source: hypercholesterol aemia See X05249 for corresponding normal gene sequence In the defective LDL-receptor gene the deletion occurred between two alu-repetitive sequences, that are in the same direction, the deletion eliminates exons 13 and 14 and changes the reading frame of the resulting spliced mRNA. Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.	
FEATURES	Location/Qualifiers 1..108 organism="Homo sapiens" /db_xref="taxon:9606" /cell_type="blood leukocytes from a patient with familial"	
Intron	1..108 /note="intron XIV fragment"	
BASE COUNT	28 a   20 c   40 g   20 t	
ORIGIN		
Query Match	'0.3%; Score 79.4; DB 10; Length 108;	
Best Local Similarity	84.8%; Pred. No. 6.1e-06;	
Matches	89; Conservative 0; Mismatches 16; Indels 0; Gaps 0;	
QY 2829	AAAAATTAGCCGGCGCTGTGGCGGCCTGTAGTCCAGCTACTTGGGAGGTGAGGC 2888   Db 3 AAAAATTAGCCAGCGTGTGGCAGGTGCTGTAACTCCAGCTACTCGGAGGTGAGGC 62	
QY 2889	AGAGAATGCGATGACCTGGGAGCGGAGCTTGCAGTAGCGCA 2933   Db 63 AGAAAATGGTTTGAACCCAGGAGCAGAGTGTGGTGAGCGCA 107	
RESULT 13	HSLDLI12 108 bp DNA PRI 20-MAY-1992	
LOCUS	Human LDL-receptor gene intron 12 fragment (normal gene) LDL = low	
DEFINITION	density lipoprotein.	
ACCESSION	X05248	
VERSION	X05248.1 GI:34334	
KEYWORDS	Alu repetitive sequence; low density lipoprotein receptor; repetitive sequence. human.	
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
REFERENCE	1 (bases 1 to 108)	
AUTHORS	Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R., Williamson,R. and Humphries,S.	
TITLE	Unequal crossing-over between two alu-repetitive DNA sequences in the low-density-lipoprotein-receptor gene. A possible mechanism for the defect in a patient with familial hypercholesterolaemia	
JOURNAL	Eur. J. Biochem. 164 (1), 77-81 (1987)	
MEDLINE	87161901	
COMMENT	*source: hypercholesterol aemia See X05248 for corresponding normal gene sequence In the defective LDL-receptor gene the deletion occurred between two alu-repetitive sequences, that are in the same direction, the deletion eliminates exons 13 and 14 and changes the reading frame of the resulting spliced mRNA. Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.	
FEATURES	Location/Qualifiers 1..108 organism="Homo sapiens" /db_xref="taxon:9606" /cell_type="blood leukocytes from a patient with familial"	
misc_feature	1..108 /note="deletion junction region Intron 12/ intron 15"	
BASE COUNT	20 a   40 C   20 g   28 t	
ORIGIN		
Query Match	0.3%; Score 79.4; DB 10; Length 108;	
Best Local Similarity	84.8%; Pred. No. 6.1e-06;	
Matches	89; Conservative 0; Mismatches 16; Indels 0; Gaps 0;	
QY 2829	AAAAATTAGCCGGCGTGTGGCGGCCTGTAGTCCAGCTACTTGGGAGGTGAGGC 2888   Db 106 AAAAATTAGCCAGCGTGTGGCAGGTGCTGTAACTCCAGCTACTCGGAGGTGAGGC 47	
QY 2889	AGAGAATGCGATGACCTGGGAGCGGAGCTTGCAGTAGCGCA 2933   Db 46 AGAAAATGGTTTGAACCCAGGAGCAGAGTGTGGTGAGCGCA 2	
RESULT 12	HSLDLR2D 108 bp DNA PRI 20-MAY-1992	
LOCUS	Human LDL-receptor mutated gene with intron 14 deletion junction.	
DEFINITION	X05251	
ACCESSION	X05251.1 GI:34336	
VERSION	Alu repetitive sequence; low density lipoprotein receptor. human.	
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;	

```
Query Match          0.3%; Score 78.2; DB 10; Length 108;
Best Local Similarity 83.2%; Pred. No. 1.1e-05;
Matches 89; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 11838 TCGGCTCACCACCACTCGCCCTCCAGGCTTCAAGCAATTCCTCGCTCAGCCTCCCA 11897
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 TCGCCTCACCACCACTCGCCTCGCTCGGTTCAAACCAATTTCTCGCTCAGCCTCCTTA 61
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 11898 GTAATGGGACTACTGGCAAGCGCCACACGCTGGCTGCTAATTTGTA 11944
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 GTAGCTGGGATTACAAGCATGTGTCACCACCGCCGGCTGATTTGTA 108
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
HSU67804
LOCUS HSU67804 108 bp RNA PRI 01-AUG-1997
DEFINITION Human small cytoplasmic Alu transcript.
ACCESSION U67804
VERSION U67804.1 GI:2289918
KEYWORDS Alu.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE cDNAs derived from primary and small cytoplasmic Alu (scAlu)
transcripts
J. Mol. Biol. 271 (2), 222-234 (1997)
MEDLINE 97415756
REFERENCE 2 (bases 1 to 108)
AUTHORS Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The
Children's Hospital of Philadelphia, 1004F Abramson Research
Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA
FEATURES
source
1..108
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TscAlu3"
1..108
/note="scAlu"
/rpt_family="Alu"
/rpt_type="dispersed"
repeat_region 26 a 38 c 26 g 18 t

BASE COUNT 26 a 38 c 26 g 18 t
ORIGIN

Query Match          0.3%; Score 77.8; DB 11; Length 108;
Best Local Similarity 87.6%; Pred. No. 1.3e-05;
Matches 85; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 9312 GCCTGTATCCAGCAGCTTTGGAGCGCCAGCGGCGGATCACGAGTCAGGAGATGGA 9371
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 GCCTGTATCCAGCAGCTTTGGAAGCGCAAGAGGAGGATCAAGGTCAGGAGATCGA 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 9372 GACCATCTGCTTAACACGATGAACCCCGTCTCTAC 9408
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GACCATCTCTGGCTAATACATGGTGAACCCCGTCTTCC 97
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
HUMALCE272
LOCUS HUMALCE272 104 bp ss-RNA PRI 15-APR-1994
DEFINITION Human carcinoma cell-derived Alu RNA transcript, clone CE272.
ACCESSION M87899
VERSION M87899.1 GI:174875
KEYWORDS Alu repeat.
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM Homo sapiens
```

```
REFERENCE 1 (bases 1 to 104)
AUTHORS Sinnott,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
post-transcriptional selection of master sequences
J. Mol. Biol. (1992) In press
FEATURES
Location/Qualifiers
1..104
/organism="Homo sapiens"
/db_xref="taxon:9606"
/dev_stage="embryo"
/sex="male"
/tissue_type="carcinoma"

BASE COUNT 22 a 26 c 37 g 19 t
ORIGIN

Query Match          0.3%; Score 76; DB 9; Length 104;
Best Local Similarity 85.0%; Pred. No. 2.9e-05;
Matches 85; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 2702 GGCCGGCGCGGTGCTCAGCCTGTATATCTTAGCAGCTTTGGAGCGCCAGACGGCGGA 2761
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5 GGCCGGCGCGGTGCTCAGCCTGTATATCTTAGCAGCAGCTTTGGAGCGCTGAGGAGGAAGA 64
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2762 TCACGAGGTCAGGAGATCGAGACCATCTTGGCTAACACGG 2801
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 TTGCGAGCGCAGGAGTTCAAGACCAGCCCTGGCTAACATGG 104
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: June 19, 2000, 03:02:32
Job time: 397436 sec
```





Result No.	Score	Query		DB	ID	Description
		Match	Length			
C 1	71.6	0.2	108	1	T26828	Human gene signatu
C 2	67.2	0.2	100	1	T24892	Human gene signatu
C 3	66.8	0.2	108	1	X12095	Human biallelic po
C 4	63.8	0.2	86	1	Y41231	Mouse embryonic ce
C 5	64	0.2	100	1	T24892	Human gene signatu
C 6	62.8	0.2	103	1	T20927	Human gene signatu
C 7	62.8	0.2	108	1	X12095	Human biallelic po
C 8	61	0.2	108	1	T25009	Human gene signatu
C 9	60	0.2	93	1	T25688	Human gene signatu
C 10	60.2	0.2	108	1	T25009	Human gene signatu
C 11	59.2	0.2	100	1	X12087	Human biallelic po
C 12	59.2	0.2	100	1	X12085	Human biallelic po
C 13	58.6	0.2	110	1	T26288	Human gene signatu
C 14	58	0.2	100	1	X12086	Human biallelic po
C 15	57.4	0.2	110	1	T25260	Human gene signatu
C 16	57	0.2	108	1	T26828	Human gene signatu
C 17	56.4	0.2	99	1	T20931	Human gene signatu
C 18	56.2	0.2	103	1	T26213	Human gene signatu
C 19	55	0.2	100	1	X12087	Human biallelic po
C 20	55	0.2	100	1	X12085	Human biallelic po
C 21	55	0.2	100	1	X12086	Human biallelic po
C 22	54.6	0.2	69	1	Q29016	Probe to internal
C 23	54.6	0.2	91	1	T25854	Human gene signatu
C 24	54.8	0.2	97	1	T26728	Human gene signatu
C 25	54.4	0.2	109	1	T23895	Human gene signatu
C 26	53.4	0.2	97	1	T26728	Human gene signatu
C 27	53.6	0.2	100	1	Q76490	Human genome fragm
C 28	53.2	0.2	82	1	T25468	Human gene signatu
C 29	53.2	0.2	110	1	T26288	Human gene signatu
C 30	51.6	0.2	89	1	T23513	Human gene signatu
C 31	51.4	0.2	91	1	T25854	Human gene signatu
C 32	51.4	0.2	103	1	T26213	Human gene signatu
C 33	50.4	0.2	62	1	T25689	Human gene signatu
C 34	50.4	0.2	102	1	T20743	Human gene signatu





PT New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease  
PS Claim 1: Page 219; 310pp; English.  
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
CC markers which have been isolated using the primers represented in  
CC X09121-X10268. The base occupying the polymorphic site is indicated by  
CC the appropriate IUPAC-TUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.  
SQ Sequence 108 BP; 19 A; 23 C; 28 G; 37 T;

Query Match 0.2%; Score 62.8; DB 1; Length 108;  
Best Local Similarity 76.0%; Pred. No. 0.045; Mismatches 23; Indels 0; Gaps 0;  
Matches 76; Conservative 1;  
Qy 7108 TTAATCCGCGCAATTTGGGAGGCGGAGCGGGATCAGTCTGAGTGGGAGTTCGAGA 7167  
Db 105 TAATCCAGCACTTTGGGAGGCGGAGCGGAGCGGATCAGTCTGAGTGGGAGTTCGAGA 46  
Qy 7168 CTAGCCCGCGCAATGCGGGAAGCCCTCTCTACTATAAA 7207  
Db 45 CCATCTCGCCCAACAYAGGAAACCTCATCTCTACAAAAA 6

## RESULT 8

ID T25009/c  
AC T25009;  
DT 07-NOV-1996 (first entry)  
DE Human gene signature HUMGS07131.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN WO9514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI: 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1: Page 1748; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 108 BP; 34 A; 31 C; 26 G; 15 T;  
Query Match 0.2%; Score 61; DB 1; Length 108;  
Best Local Similarity 74.5%; Pred. No. 0.089; Mismatches 26; Indels 0; Gaps 0;  
Matches 76; Conservative 0;  
Qy 10694 TTTTGTGTTTTCAGACTGAGCTTGTCTGTCCACCCAGCTGGAGTGAATGGCGC 10753  
Db 103 TGTGTTGTTCTTTTCAACAGGGTCTTGTCTGTCTGTCACATCAGTGGTGGCGTG 44  
Qy 10754 ATCTCGGCTCACTCAACCTCCGCTTCCAGGTTCACGGAT 10795  
Db 43 ACCATGGCTCACTGAGCGCTTGGGCTCATGGGCTCAGGCGAT 2

## RESULT 9

ID T25888/c  
AC T25888;  
DT 09-OCT-1996 (first entry)  
DE Human gene signature HUMGS07887.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN WO9514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI: 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1: Page 1907; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 93 BP; 25 A; 27 C; 24 G; 17 T;  
Query Match 0.2%; Score 60; DB 1; Length 93;  
Best Local Similarity 78.3%; Pred. No. 0.13; Mismatches 20; Indels 0; Gaps 0;  
Matches 72; Conservative 0;  
Qy 13277 TTTGAGATGGAATTTCACTCTTGTGTCACCGCTGGGTGGCAGTGCACAGTTCAGCTC 13336  
Db 93 TTTGAGATGGGTCTCTCACTCTTGTCAACCCAGCTGAGTGGTGGTCAATCAGAGTTC 34  
Qy 13337 ACTGCAACCTCCGCTCTCTGGGTTCACGGAT 13368  
Db 33 ACTGTGGCTTCGGCTTCTCTGGGCTCAAGAGAT 2



```
RESULT 10
T25009 ID T25009 standard; cDNA to mRNA; 108 BP.
AC T25009;
DE 07-NOV-1996 (first entry)
DE Human gene signature HUMGS07131.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW cell; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues.
PS Claim 1; Page 1748; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 108 BP; 34 A; 31 C; 26 G; 15 T;

Query Match 0.2%; Score 60.2; DB 1; Length 108;
Best Local Similarity 72.0%; Pred. No. 0.12;
Matches 77; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 9486 ATGGCGTGAACGGGGGAGCGGAGCTTGCAGTGAGCGGAGATCGCGCCAGGCGACTCCAG 9545
Db 2 ATCGCCTGAGCCCATGATGAGGCCAAGGCTGCAGTGAGCGATGTCACGCCACTGNATTCAG 61

QY 9546 CCTGGGTGACAGCGGAGAGCTCCGTCTCAAAAAAAAAAAAAAAAAAAAAA 9592
Db 62 CCTGAGTGACAGACGACAGACCCCTGTTGAAAAACACACACACANCAA 108

RESULT 11
X12087/c ID X12087 standard; DNA; 100 BP.
AC X12087;
DE 30-MAR-1999 (first entry)
DE Human biallelic polymorphic DNA fragment EST98276a.
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW treatment; marker; ss.
OS Homo sapiens.
PN WO9820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; U20313.
PR 06-NOV-1996; US-030455.
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
DR WPI; 98-286974/25.

PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
PS Claim 1; Page 219; 310pp; English.
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC X09121-X10268. The base occupying the polymorphic site is indicated by
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC segments can also be used to produce medicaments for the treatment or
CC prophylaxis of such diseases.
SQ Sequence 100 BP; 21 A; 25 C; 22 G; 31 T;

Query Match 0.2%; Score 59.2; DB 1; Length 100;
Best Local Similarity 74.5%; Pred. No. 0.17;
Matches 73; Conservative 1; Mismatches 24; Indels 0; Gaps 0;

QY 8434 TGCAGCTCAGCCCTGTAATACACAGCATTTTGGGAGGCCAAGGTGGGAGATCACTTGCAGC 8493
Db 100 TGTGACTCACACCTTAATCTCGCACTTTAGGAGGCTTAGGAGGAGGATGTTTGAAGA 41

QY 8494 CCAGGAGCTCAAGACCACTCTGGGCAACTTAGTGAGAC 8531
Db 40 CCAGGAGCTCAAGACCAKCTCTGGGAAACATAGCAAGAC 3

RESULT 12
X12085/c ID X12085 standard; DNA; 100 BP.
AC X12085;
DE 30-MAR-1999 (first entry)
DE Human biallelic polymorphic DNA fragment EST98276c.
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW treatment; marker; ss.
OS Homo sapiens.
PN WO9820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; U20313.
PR 06-NOV-1996; US-030455.
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
DR WPI; 98-286974/25.

PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
PS Claim 1; Page 218; 310pp; English.
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC X09121-X10268. The base occupying the polymorphic site is indicated by
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
```

CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.  
SQ Sequence 100 BP; 22 A; 25 C; 22 G; 30 T;

Query Match 0.2%; Score 59.2; DB 1; Length 100;  
Best Local Similarity 74.5%; Pred. No. 0.17;  
Matches 73; Conservative 1; Mismatches 24; Indels 0; Gaps 0;

QY 8434 TGCAGCTCAGGCTGTATACACGACATTTGGGAGGCCAAGGTGGGAGGATCATTGAGC 8493  
II IIIII III IIII III IIII IIII IIII IIII IIII IIII IIII  
DB 100 TGTGACTCACCTATATCTCTGGCACTTTGGAGGCTTAGGAGGAGGATTTGTTGAA 41

QY 8494 CCAGGAGCTCAAGACCACTCTGGGCACTTAGTGAGAC 8531  
IIIIII IIIII IIII IIII IIII IIII IIII IIII IIII IIII  
DB 40 CCAGGAGCTCAAGACCACTCTGGGAAACATAGCAAGAC 3

RESULT 13  
T26288  
ID T26288 standard; cDNA to mRNA; 110 BP.  
AC T26288;  
DE 10-OCT-1996 (first entry)  
DT Human gene signature HUMG808527.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN WO9514772-Al.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATSU) MATSUBARA K.  
PA (OKUBU) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI; 95-206931/27  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1; Page 2048; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-R26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 110 BP; 22 A; 37 C; 28 G; 17 T;

Query Match 0.2%; Score 58.6; DB 1; Length 110;  
Best Local Similarity 85.3%; Pred. No. 0.22;  
Matches 64; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 12008 GATCCACCGGCTCGGCTCCCAAGTCTGGGATTCAGAGTATGAGCACTGGGCCCGG 12087  
IIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII  
DB 1 GATCCGCGGCTCGAGCTCCCAAGTCTGGGATTCAGAGTATGAGCACTGGCACCGG 60

QY 12068 CCACATTTCTAAAT 12082

DB 61 CCCCATTCCTCACTT 75  
II IIIII III IIII  
RESULT 14  
X12086/c  
ID X12086 standard; DNA; 100 BP.  
AC X12086;  
DE 30-MAR-1999 (first entry)  
DT Human biallelic polymorphic DNA fragment EST98276b.  
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;  
KW detection; phenotypic typing; characteristic; infection; hereditary;  
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;  
KW treatment; marker; ss.  
OS Homo sapiens.  
PN WO9820165-A2.  
PD 14-MAY-1998.  
PF 05-NOV-1997; U20313.  
PR 06-NOV-1996; US-030455.  
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.  
PI Hudson T, Lander ES, Wang D;  
DR WPI; 98-286974/25.  
PT New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease  
PS Claim 1; Page 219; 310pp; English.  
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
CC markers which have been isolated using the primers represented in  
CC X09121-X10268. The base occupying the polymorphic site is indicated by  
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as acamaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberculous scleriosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.  
SQ Sequence 100 BP; 21 A; 25 C; 22 G; 31 T;

Query Match 0.2%; Score 58; DB 1; Length 100;  
Best Local Similarity 74.5%; Pred. No. 0.27;  
Matches 73; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 8434 TGCAGCTCAGGCTGTATACACGACATTTGGGAGGCCAAGGTGGGAGGATCATTGAGC 8493  
II IIIII III IIII III IIII IIII IIII IIII IIII IIII  
DB 100 TGTGACTCACCTATATCTCTGGCACTTTAGGAGGCTTAGGAGGAGGATTTGTTGAA 41

QY 8494 CCAGGAGCTCAAGACCACTCTGGGCACTTAGTGAGAC 8531  
IIIIII IIIII IIII IIII IIII IIII IIII IIII IIII IIII  
DB 40 CCAGGAGCTCAAGACCACTCTGGGAAACATAGCAAGAC 3

RESULT 15  
T25260/c  
ID T25260 standard; DNA; 110 BP.  
AC T25260;  
DE 11-NOV-1996 (first entry)  
DT Human gene signature HUMG507421.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN WO9514772-Al.  
PD 01-JUN-1995.





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2000, 12:40:45 ; Search time 8475.2 Seconds  
(without alignments)  
13869.588 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_84000\_113000  
Perfect score: 29001  
Sequence: 1 TCAAACTCCTGACCTCATGA.....ATAAATATCTTAAATAACC 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 156056

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result % Query

No.	Score	Match	Length	DB	ID	Description
C 1	96.2	0.3	109	30	AA243009	zr25h02.s
C 2	94.8	0.3	106	37	AA703692	ag81a10.r
C 3	93.8	0.3	105	61	AI832832	at832832 at72g09.x
C 4	93.2	0.3	106	38	AA812141	ob48h02.s
C 5	91.8	0.3	103	84	B48914	RPC111-4A12
C 6	91.4	0.3	101	39	AA835205	ak64h01.s
C 7	91.6	0.3	109	24	N25299	yw52c09.s1
C 8	91.4	0.3	109	30	AA243009	zr25h02.s
C 9	91.6	0.3	110	39	AA897366	am06h02.s
C 10	91.2	0.3	107	39	AA828124	od71a07.s
C 11	89.6	0.3	97	25	N49638	yv25e09.r1
C 12	88.2	0.3	103	94	AQ028649	CIT-HSP-2
C 13	88.2	0.3	101	94	AQ076649	CIT-HSP-2
C 14	88.2	0.3	102	30	AA226656	nc19f09.s
C 15	88.4	0.3	106	105	AQ264176	CITBI-E1-
C 16	88	0.3	107	33	AA385808	EST99495
C 17	87.6	0.3	103	38	AA807640	nx08b05.s
C 18	87.6	0.3	110	79	AA250394	2822460.3
C 19	86.8	0.3	106	37	AA703692	ag81a10.r
C 20	85.8	0.3	106	108	AQ544457	CITBI-E1-
C 21	85.8	0.3	107	24	H67040	yu68c01.r1
C 22	86	0.3	110	64	AW083640	xc49f02.x
C 23	85.4	0.3	103	84	B48914	RPC111-4A12
C 24	85	0.3	109	84	B17434	345K2.TVB C
C 25	85	0.3	110	94	AQ003188	RPC111-1D
C 26	84.4	0.3	103	108	AQ582186	RPC111-4
C 27	83.8	0.3	103	30	AA228795	nc14e07.s
C 28	83.8	0.3	105	109	AQ637292	RPC111-4
C 29	84	0.3	110	33	AA442529	zv68b02.r
C 30	83.6	0.3	106	34	AA516339	ng71g02.s
C 31	83.6	0.3	107	35	AA565533	nk42b11.s
C 32	83.2	0.3	107	24	N23686	yw46a02.s1
C 33	83.4	0.3	109	94	AQ028426	CIT-HSP-2
C 34	83.4	0.3	110	94	AQ003188	RPC111-1D
C 35	82.8	0.3	102	36	AA654562	nt75f10.s
C 36	82.8	0.3	105	74	AW196212	xm06e06.x
C 37	83	0.3	107	62	AI933497	wm74d02.x
C 38	82.4	0.3	93	43	AI158167	oc09e10.x
C 39	82.6	0.3	108	32	AA370029	EST81584
C 40	82	0.3	100	35	AA564832	nj22a06.s
C 41	82.2	0.3	103	94	AQ028649	CIT-HSP-2
C 42	82.2	0.3	103	108	AQ584425	RPC111-4
C 43	81.8	0.3	101	35	AA583697	nn58f10.s
C 44	82	0.3	106	38	AA812141	ob48h02.s
C 45	82	0.3	106	94	AQ062963	CIT-HSP-2

## ALIGNMENTS

```

RESULT 1
AA243009/c 109 bp mRNA EST 11-MAR-1998
LOCUS zr25h02.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens
DEFINITION cDNA clone IMAGE:664467 3' similar to contains Alu repetitive
            element;contains element LTR1 repetitive element ;, mRNA sequence.

ACCESSION AA243009
VERSION   AA243009.1
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 109)
AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
           Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
           Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
           Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
           WashU-NCI human EST Project
           Unpublished (1997)
TITLE     WashU-NCI human EST Project
JOURNAL
COMMENT   On Dec 3, 1996 this sequence version replaced gi:1126869.

```

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1127 Std Error: 0.00  
Seq primer: -41m13 fwd. ST from Amersham  
High quality sequence stop: 102.

## FEATURES

Location/Qualifiers  
1..109  
/organism="Homo sapiens"  
/db\_xref="GDB:5426481"  
/db\_xref="taxon:9606"  
/clone="IMAGE:664467"  
/clone\_lib="Stratagene NT2 neuronal precursor 937230"  
/tissue\_type="neuroepithelial cells"  
/dev\_stage="Ntera-2 neuroepithelial cells"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: brain; Vector: pBluescript SK-; Site:1:  
EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Uninduced, exponentially growing neuroepithelial  
cells (Ntera-2/ci.D1). Average insert size: 1.0 kb;  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG  
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3' "

BASE COUNT 19 a 30 c 30 g 30 t

## ORIGIN

Query Match 0.3%; Score 96.2; DB 30; Length 109;  
Best Local Similarity 92.7%; Pred. No. 0.076; 8; Indels 0; Gaps 0;  
Matches 101; Conservative 0; Mismatches 0;  
QY 9309 CACGCTTAATCCAGCACCTTTGGAGCGCGGATCAGGATCAGGATCAGGATCAGGAT 9368  
DB 109 CACGCTTAATCCAGCACCTTTGGAGCGCGGATCAGGATCAGGATCAGGATCAGGAT 50  
QY 9369 GGAGACCATCTCTGTACACGATGAAACCCCGTCTCTACTAAAAATAC 9417  
DB 49 CAAGACCATCTCTGTACACGATGAAACCCCGTCTCTACTAAAAATAC 1

## RESULT 2

AA703692/c 106 bp mRNA EST 24-DEC-1997  
LOCUS ag81a10.r1 Stratagene hnt neuron (#937233) Homo sapiens cDNA clone  
DEFINITION IMAGE:1140858 5' similar to contains Alu repetitive element; , mRNA  
sequence.

ACCESSION AA703692

VERSION AA703692.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 106)

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,

Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,

Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,

Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.

WashU-NCI human EST Project

Unpublished (1997)

On Sep 12, 1996 this sequence version replaced gi:1397630.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 53.

# FEATURES

source

1. .106  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1140858"  
/clone\_lib="Stratagene hMT neuron (#937233)"  
/dev\_stage="hMT neurons"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Differntiated, post mitotic hMT neurons. Average insert size: 1.5 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCAGAG 3' ~3' adaptor sequence: 5' CTCGAGTCTTTTCTTTT 3' "

BASE COUNT 19 a 29 c 29 g 29 t  
ORIGIN

Query Match 0.3%; Score 94.8; DB 37; Length 106;  
Best Local Similarity 93.4%; Pred. No. 0.11; 7; Indels 0; Gaps 0;  
Matches 99; Conservative 0; Mismatches 0; Gaps 0;

QY 9309 CACGCTGTAAATCCAGCACTTTGGGAGCGCGGATCAGGATCAGGAT 9368  
Db 106 CACGCTGTAAATCCAGCACTTTGGGAGCGCGGATCAGGATCAGGAT 47  
QY 9369 GGAGACCATCTCGTTAAGACAGTAAGACCCGCTCTCTACTAAAA 9414  
Db 46 CGAGACCATCTCGTTAAGACAGTAAGACCCGCTCTCTACTAAAA 1

# RESULT 3

AI832832 105 bp mRNA EST 13-JUL-1999  
LOCUS at72g09.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone  
DEFINITION IMAGE:2377600 3' similar to contains Alu repetitive  
element;contains element MER22 repetitive element ;, mRNA sequence.  
ACCESSION AI832832  
VERSION AI832832.1 GI:5454812  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 105)  
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,  
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,  
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
TITLE WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT On Dec 20, 1995 this sequence version replaced gi:1133644.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from Gibco.

# FEATURES

source

1. .105  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2377600"  
/clone\_lib="Barstead colon HPLRB7"  
/sex="male"  
/dev\_stage="adult, age 25"  
/lab\_host="DH10B (phage resistant)"  
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a

modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st  
strand cDNA was primed with a Not I - oligo(dt) primer [5'  
TGTTACGATCTGAAGTGGAGCGGCCCTTTTCTTTTCTTTTCTTTTCTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
[5' AATTCACCTAGTAAT 3' and 5' ATTACTAGTG 3'], digested  
with Not I and cloned into the Not I and Eco RI sites of  
the modified pT7T3 vector. Library constructed by Bob  
Barstead."

BASE COUNT 17 a 35 c 27 g 26 t  
ORIGIN

Query Match 0.3%; Score 93.8; DB 61; Length 105;  
Best Local Similarity 93.3%; Pred. No. 0.14; 7; Indels 0; Gaps 0;  
Matches 98; Conservative 0; Mismatches 0; Gaps 0;

QY 11788 GAGACGGAGTTTCACACTTGTGCCAGGCTGGAGTGAATGTCGATCTCGCTCACC 11847  
Db 1 GAGACAGAGTTTCGCTCTTGTTCGCCAGGCTGGAGTGAATGTCGCTCACC 60  
QY 11848 GCAACCTCCGCTCCAGGTTCAAGCAATTCCTCCTCAGCCT 11892  
Db 61 GCAACCTCCAGCTCCCGGGTTCAAGCGATTCTCCTCCTCAGCCT 105

# RESULT 4

AA812141/c 106 bp mRNA EST 19-FEB-1998  
LOCUS ob48h02.sl NCI-CGAP-CCB1 Homo sapiens cDNA clone IMAGE:1334643 3'  
DEFINITION similar to contains Alu repetitive element;; mRNA sequence.  
ACCESSION AA812141  
VERSION AA812141.1 GI:2881752  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 106)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1402063.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
Ph.D., Gerald Marti, M.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/dbrrp/image/image.html

Insert Length: 1450 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 60.

# FEATURES

source

1. .106  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1334643"  
/clone\_lib="NCI-CGAP-CCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),  
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was

primed with a Not I - oligo(dT) primer  
 [5'-TGTTCAATCTCAAGTGGGCGGCGCTCATTTTTTTTTTTTTTTT-  
 3']. Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT773 vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 16 a 31 c 24 g 35 t

ORIGIN

Query Match 0.3%; Score 93.2; DB 38; Length 106;  
 Best Local Similarity 92.5%; Pred. No. 0.17;  
 Matches 98; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 9481 GGAGATGCGTGAACGGCGGAGCGGAGCTTGCAGTCGAGCGGATCGCGCCACGGCAC 9540

Db 106 GGAGATGCGTGAACCGTGGAGGTTGGAGCTTGCAGTCGAGCGGATCGACACCACTGCAC 47

QY 9541 TCCAGCGCTGGTGCACAGCGAGACTCCGCTCTCAAAAAA 9586

Db 46 TCCAGCGCTGGTGCACAGCGAGACTCCATCTCAAAAAA 1

RESULT 5

LOCUS B48914

DEFINITION B48914 103 bp DNA GSS 08-APR-1999  
 RPC111-4A12.TP RPC1-11 Homo sapiens genomic clone RPC1-11-4A12,  
 genomic survey sequence.

ACCESSION B48914

VERSION B48914.1 GI:2601151

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 103)

AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,  
 Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and  
 Venter,J.C.

TITLE Use of BAC End Sequences for Sequence-Ready Map Building

JOURNAL Unpublished (1997)

COMMENT Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are derived from the human BAC library RPC1-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from

Research Genetics (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/hungen/bac\_end\_search/bac\_end\_search.html

seq primer: SP6

Class: BAC ends.

FEATURES

source

1..103

/organism="Homo sapiens"

/db\_xref="GDB:7501163"

/db\_xref="taxon:9606"

/clone="RPC1-11-4A12"

/clone\_lib="RPC1-11"

/sex="Male"

/cell\_type="Lymphocytes"

/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;

RPC111 Human Male BAC Library"

BASE COUNT 30 a 28 c 30 g 15 t

ORIGIN

Query Match 0.3%; Score 91.8; DB 84; Length 103;

Best Local Similarity 93.2%; Pred. No. 0.24;  
 Matches 96; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 9316 GTAATCCAGCACTTTGGAGGCGGAGCGGATCACGAGGTCAGGAGATCGAGACC 9375

Db 1 GTAAGCCAGCACTTTGGAGGCGGAGCGGATCACGAGGTCAGGAGATCGAGACC 60

QY 9376 ATCTCGCTTAACACGATGAACCCCGCTCTACTACTAAAAATACA 9418

Db 61 ATCCGCGCTAAACGGTGAACCCCGCTCTACTAAAAATACA 103

RESULT 6

LOCUS AA835205/c

DEFINITION at64h01.sl Barstead pancreas HPLRB1 Homo sapiens cDNA clone

IMAGE:1412689 3' similar to contains Alu repetitive

element;contains element KER repetitive element ;, mRNA sequence.

ACCESSION AA835205

VERSION AA835205.1 GI:2908933

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 101)

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,

Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,

Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,

Theising,B., White,Y., Wyllie,F., Waterston,R. and Wilson,R.

WashU-NCI human EST Project

Unpublished (1997)

On Nov 29, 1993 this sequence version replaced gi:636191.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES

source

Location/Qualifiers

1..101

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1412689"

/clone\_lib="Barstead pancreas HPLRB1"

/sex="female"

/dev\_stage="adult, 34 years"

/lab\_host="DH10B"

/note="Organ: pancreas; Vector: pT7T3D-Pac (Pharmacia)

with a modified polylinker; Site\_1: EcoRI; Site\_2: NotI;

1st strand cDNA was primed with a Not I - oligo(dT) primer

[5'

TGTTACGAATCGAAGTGGGAGCGCGCCCTTTTTTTTTTTTTTTTTT

3']; double-stranded cDNA was ligated to Eco RI adaptors

[AATTCGATCCTTG], digested with Not I and cloned into the

Not I and Eco RI sites of the modified pT7T3 vector.

Library constructed by Bob Barstead."

BASE COUNT 14 a 36 c 27 g 24 t

ORIGIN

Query Match 0.3%; Score 91.4; DB 39; Length 101;

Best Local Similarity 94.1%; Pred. No. 0.27;  
 Matches 95; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 9474 TGAGGAGAGAAATGCGTGAACGGCGGAGCGGAGCTTGCAGTGCAGGCGAGATCGCGCC 9533

Db 101 TGAGGAGAGAAATGCGTGAACCGCGGAGCGGAGCTTGCAGTGCAGGCGAGATCAAGCC 42

QY 9534 ACGGCACCTCCAGCCTGGGTGACAGAGCGAGACTCCGCTCTCA 9574



```

Db 41 ACTGCACTCCAGCCTGGGCGACAGAGTGCAGCTCCGCTCA 1
|| |||||
RESULT 7
N25299/c 109 bp mRNA EST 28-DEC-1995
LOCUS yw52c09.s1 Weizmann Olfactory Epithelium Homo sapiens cDNA clone
DEFINITION IMAGE:255856 3' similar to contains Alu repetitive element;; mRNA
sequence.
ACCESSION N25299
VERSION N25299.1 GI:1139449
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 109)
AUTHORS Hillier,L., Lannon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT On Apr 14, 1993 this sequence version replaced gi:837394.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: m13 -40 forward
High quality sequence stop: 307.
FEATURES
Location/Qualifiers
1..109
source
/organism="Homo sapiens"
/db_xref="GB:3866265"
/db_xref="taxon:9606"
/clone="IMAGE:255856"
/clone_lib="Weizmann Olfactory Epithelium"
/sex="Female"
/tissue_type="olfactory epithelium"
/dev_host="35 year old"
/label="SOLR cells (kanamycin resistant)"
/note="Organ: nose; Vector: pBluescript SR-; Site_1:
ECORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Olfactory epithelium, normal. Average insert
size: 0.8 kb; Uni-ZAP XR Vector. Library constructed by N.
Walker, D. Lancet, Weizmann Institute of Science. -5'
adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTT 3'
BASE COUNT 13 a 34 c 24 g 35 t 3 others
ORIGIN
Query Match 0.3%; Score 91.6; DB 24; Length 109;
Best Local Similarity 89.0%; Pred. No. 0.24; Indels 0; Gaps 0;
Matches 97; Conservative 0; Mismatches 12;
QY 2885 AGCAGGAGAAATGGCACTGGGAGCGGAGCTTGCAGTGCAGCGCGCCAC 2944
Db 109 AGCAGGAGAAATGGCACTGGGAGCGGAGCTTGCAGTGCAGCGCGCCAC 50
QY 2945 TGCACCTCCAACTGGGAGACACGCGAGACTCGGTCTCAAAAAA 2993
Db 49 TGCACCTCCAGCCTGGGAGACACGCGAGCTCGGTCTCAAAAAA 1

```

---

```

RESULT 8
AA243009 109 bp mRNA EST 11-MAR-1998
LOCUS zr23h02.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens
DEFINITION cDNA clone IMAGE:664467 3' similar to contains Alu repetitive
element; contains element Ltr1 repetitive element;; mRNA sequence.
ACCESSION AA243009
VERSION AA243009.1 GI:1873869
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 109)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krisman,D., Kucaba,T., Lacy,M., Le,N., Lannon,G., Marra,M.,
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
JOURNAL On Dec 3, 1996 this sequence version replaced gi:1126869.
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1127 Std Error: 0.00
Seq primer: -41ml3 fwd. Et from Amersham
High quality sequence stop: 102.
FEATURES
Location/Qualifiers
1..109
source
/organism="Homo sapiens"
/db_xref="GB:5426481"
/db_xref="taxon:9606"
/clone="IMAGE:664467"
/clone_lib="Stratagene NT2 neuronal precursor 937230"
/tissue_type="neuroepithelial cells"
/dev_host="Ntera-2 neuroepithelial cells"
/label="SOLR (kanamycin resistant)"
/note="Organ: brain; Vector: pBluescript SK-; Site_1:
ECORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Uninduced, exponentially growing neuroepithelial
cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3'
BASE COUNT 19 a 30 c 30 g 30 t
ORIGIN
Query Match 0.3%; Score 91.4; DB 30; Length 109;
Best Local Similarity 89.9%; Pred. No. 0.26;
Matches 98; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 10862 GTATTTTACTAGACAGAGGTTTCACCGTGTGGCGCAGGATGTCFAATCTCCTTACC 10921
Db 1 GTATTTTACTAGACAGCGGGTTTCACCGTGTGGCGCAGGATGTCCTGATCTCCCTACC 60
QY 10922 TCCTGATCCCGCCCTCGCTCCGCAAGTCTCGGATTACAGACGTG 10970
Db 61 TCCTGATCCCGCCCTCGCTCCGCAAGTCTCGGATTACAGACGTG 109
RESULT 9
AA897366/c 110 bp mRNA EST 04-JAN-1999
LOCUS am06h02.s1 Soares_NFL_T_GBC.s1 Homo sapiens cDNA clone
DEFINITION IMAGE:1466067 3' similar to contains Alu repetitive element;; mRNA
sequence.

```

AA897366  
AA897366.1 GI:3033986  
EST.  
SOURCE  
ORGANISM  
Human sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 110)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Jan 19, 1998 this sequence version replaced gi:2150764.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
Insert Length: 834 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 63.  
Location/Qualifiers  
1..110  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:I466067"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT7m3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19w, testis NHT, and B-cell  
NCI-CGAP GCB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
736408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
22 a 27 c 29 g 32 t  
BASE COUNT  
ORIGIN  
Query Match 0.3%; Score 91.6; DB 39; Length 110;  
Best Local Similarity 91.5%; Pred. No. 0.24;  
Matches 97; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 9309 CAGCGCTGAATCCAGCACTTTGGAGGCGCGGCGGATCAGGAGAT 9368  
|||||  
Db 110 CAGCGCTAATCCAGCACTTTGGAGGCGCGGCGGATCAGGAGAT 51  
QY 9369 GGAGACCATCTCTTAACACGATGAACCCCGTCTCTACTAAAA 9414  
|||||  
Db 50 TGAGACCATCTCTGTTAACACGATGAACCCCGTCTCTACTAAAA 5  
RESULT 10  
AA828124  
LOCUS  
AA828124 107 bp mRNA EST 20-FEB-1998  
DEFINITION  
od71a07.s1 NCI-CGAP\_Ov2 Homo sapiens cDNA clone IMAGE:1373364  
similar to contains Alu repetitive element; contains element MER22  
repetitive element ; mRNA sequence.  
AA828124  
ACCESSION  
AA828124.1 GI:2900487  
VERSION  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 107)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Jan 19, 1998 this sequence version replaced gi:2150764.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
Insert Length: 834 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 63.  
Location/Qualifiers  
1..110  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:I466067"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT7m3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19w, testis NHT, and B-cell  
NCI-CGAP GCB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
736408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
22 a 27 c 29 g 32 t  
BASE COUNT  
ORIGIN  
Query Match 0.3%; Score 91.6; DB 39; Length 110;  
Best Local Similarity 91.5%; Pred. No. 0.24;  
Matches 97; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 9309 CAGCGCTGAATCCAGCACTTTGGAGGCGCGGCGGATCAGGAGAT 9368  
|||||  
Db 110 CAGCGCTAATCCAGCACTTTGGAGGCGCGGCGGATCAGGAGAT 51  
QY 9369 GGAGACCATCTCTTAACACGATGAACCCCGTCTCTACTAAAA 9414  
|||||  
Db 50 TGAGACCATCTCTGTTAACACGATGAACCCCGTCTCTACTAAAA 5  
RESULT 10  
AA828124  
LOCUS  
AA828124 107 bp mRNA EST 20-FEB-1998  
DEFINITION  
od71a07.s1 NCI-CGAP\_Ov2 Homo sapiens cDNA clone IMAGE:1373364  
similar to contains Alu repetitive element; contains element MER22  
repetitive element ; mRNA sequence.  
AA828124  
ACCESSION  
AA828124.1 GI:2900487  
VERSION  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 107)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Jan 19, 1998 this sequence version replaced gi:2150764.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
Insert Length: 834 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 63.  
Location/Qualifiers  
1..110  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:I466067"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT7m3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19w, testis NHT, and B-cell  
NCI-CGAP GCB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
736408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
22 a 27 c 29 g 32 t  
BASE COUNT  
ORIGIN  
Query Match 0.3%; Score 91.6; DB 39; Length 110;  
Best Local Similarity 91.5%; Pred. No. 0.24;  
Matches 97; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 9309 CAGCGCTGAATCCAGCACTTTGGAGGCGCGGCGGATCAGGAGAT 9368  
|||||  
Db 110 CAGCGCTAATCCAGCACTTTGGAGGCGCGGCGGATCAGGAGAT 51  
QY 9369 GGAGACCATCTCTTAACACGATGAACCCCGTCTCTACTAAAA 9414  
|||||  
Db 50 TGAGACCATCTCTGTTAACACGATGAACCCCGTCTCTACTAAAA 5  
RESULT 10  
AA828124  
LOCUS  
AA828124 107 bp mRNA EST 20-FEB-1998  
DEFINITION  
od71a07.s1 NCI-CGAP\_Ov2 Homo sapiens cDNA clone IMAGE:1373364  
similar to contains Alu repetitive element; contains element MER22  
repetitive element ; mRNA sequence.  
AA828124  
ACCESSION  
AA828124.1 GI:2900487  
VERSION  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 107)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Jan 17, 1998 this sequence version replaced gi:1899815.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
[www-bio.lnl.gov/bbrp/image/image.html](http://www-bio.lnl.gov/bbrp/image/image.html)  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 93.  
Location/Qualifiers  
1..107  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1373364"  
/clone\_lib="NCI-CGAP\_Ov2"  
/sex="female"  
/tissue\_type="ovary"  
/lab\_host="DH10B"  
/note="Vector: PAM10: mRNA made from invasive ovarian  
tumor, cDNA made by oligo-dT priming. Non-directionally  
cloned. Size-selected on agarose gel, average insert size  
600 bp. Reference: Krizman et al. (1996) Cancer Research  
56:5380-5383."  
30 a 23 c 38 g 16 t  
BASE COUNT  
ORIGIN  
Query Match 0.3%; Score 91.2; DB 39; Length 107;  
Best Local Similarity 92.3%; Pred. No. 0.27;  
Matches 96; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 2814 TCTACTAAATAATACAAAAATTTAGCGGGCGGTGTGGCGGCGCGCTAGTCCAGCTAC 2873  
|||||  
Db 4 TCGACTAAATAATACAAAAATTTAGCGGGCGGTGTGGCGGCGCGCTAGTCCAGCTAC 63  
QY 2874 TTGGAGGCTGAGGAGGAGATGCGCATGAACCTGGGAGGCGGA 2917  
|||||  
Db 64 TTGGAGGCTGAGGAGGAGATGCGCATGAACCTGGGAGGCGGA 107  
RESULT 11  
N49638  
LOCUS  
N49638 97 bp mRNA EST 14-FEB-1996  
DEFINITION  
yv25e09.r1 Soares fetal liver spleen 1NF1S Homo sapiens cDNA clone  
IMAGE:243784 5' similar to gb:X57138\_rnal HISTONE H2B





Search completed: June 18, 2000, 22:02:37  
Job time: 379524 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2000, 17:46:45 ; Search time 373.09 Seconds  
(without alignments)  
10103.987 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_84000\_113000  
Perfect score: 29001  
Sequence: 1 TCAAACTCCTGACCTCATGA.....ATAAATATCTTTAAATACC 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 374504

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2.6/ptodata/1/ina/5C\_COMB.seq.\*  
4: /cgn2.6/ptodata/1/ina/5D\_COMB.seq.\*  
5: /cgn2.6/ptodata/1/ina/6\_COMB.seq.\*  
6: /cgn2.6/ptodata/1/ina/PTUS\_COMB.seq.\*  
7: /cgn2.6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80	0.3	105	4	US-08-481-658B-65
2	80	0.3	105	4	US-08-477-504A-65
3	80	0.3	105	4	US-08-486-756A-65
4	80	0.3	105	4	US-08-485-862B-65
5	80	0.3	105	5	US-08-787-739-65
6	76.2	0.3	105	4	US-08-481-658B-65
7	76.2	0.3	105	4	US-08-477-504A-65
8	76.2	0.3	105	4	US-08-486-756A-65
9	76.2	0.3	105	4	US-08-485-862B-65
10	76.2	0.3	105	5	US-08-787-739-65
11	67.6	0.2	84	3	US-08-454-557C-91
12	67.6	0.2	84	4	US-08-340-426B-91
13	67.6	0.2	84	4	US-08-450-673C-91
14	67.6	0.2	84	6	PCT-US95-17111A-91
15	62.8	0.2	76	3	US-08-454-557C-69
16	62.8	0.2	76	4	US-08-340-426B-69
17	62.8	0.2	76	4	US-08-450-673C-69
18	62.8	0.2	76	6	PCT-US95-17111A-69
19	59.2	0.2	85	3	US-08-454-557C-92
20	59.2	0.2	85	4	US-08-340-426B-92
21	59.2	0.2	85	4	US-08-450-673C-92
22	59.2	0.2	85	6	PCT-US95-17111A-92
23	58.4	0.2	60	3	US-08-454-557C-57
24	58.4	0.2	60	4	US-08-340-426B-57
25	58.4	0.2	60	4	US-08-450-673C-57
26	58.4	0.2	60	6	PCT-US95-17111A-57
27	57.4	0.2	84	3	US-08-454-557C-91

28	57.4	0.2	84	4	US-08-340-426D-91	Sequence 91, Appl
29	57.4	0.2	84	4	US-08-450-673C-91	Sequence 91, Appl
30	57.4	0.2	84	6	PCT-US95-17111A-91	Sequence 91, Appl
31	55	0.2	85	3	US-08-454-557C-92	Sequence 92, Appl
32	55	0.2	85	4	US-08-340-426B-92	Sequence 92, Appl
33	55	0.2	85	4	US-08-450-673C-92	Sequence 92, Appl
34	55	0.2	85	6	PCT-US95-17111A-92	Sequence 92, Appl
35	52	0.2	60	3	US-08-454-557C-60	Sequence 60, Appl
36	52	0.2	60	4	US-08-340-426B-60	Sequence 60, Appl
37	52	0.2	60	4	US-08-450-673C-60	Sequence 60, Appl
38	52	0.2	60	6	PCT-US95-17111A-60	Sequence 60, Appl
39	52.2	0.2	94	5	US-08-750-064-6	Sequence 6, Appl
40	50.8	0.2	78	3	US-08-454-557C-70	Sequence 70, Appl
41	50.8	0.2	78	4	US-08-340-426B-70	Sequence 70, Appl
42	50.8	0.2	78	4	US-08-450-673C-70	Sequence 70, Appl
43	50.8	0.2	78	6	PCT-US95-17111A-70	Sequence 70, Appl
44	50.4	0.2	60	5	US-08-545-860B-71	Sequence 71, Appl
45	50.4	0.2	60	6	PCT-US94-04496-71	Sequence 71, Appl

ALIGNMENTS

RESULT 1  
US-08-481-658B-65  
; Sequence 65, Application US/08481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,658B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.3E  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; US-08-481-658B-65

Query Match 0.3%; Score 80; DB 4; Length 105;  
Best Local Similarity 85.6%; Pred. No. 9.5e-10;

Matches 89; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 10857 TTTTGTATTTTAGTAGACAGGTTTCACCGTGTGGCCAGGATGTTCTCAATCTCC 10916

Db 2 TTTTACATCTTTAGTAGACAGGTTTCACCATATTGGCCAGGCTGCTCTCAAACTCC 61

QY 10917 TTACCTCGTGATCCGCCCTCGCTGCTGCCAAAGTGTCTGGAT 10960

Db 62 TGACCTGTGTGATCCACGAGCTCGGCTCCCAAAAGTGTGGAT 105

## RESULT 2

US-08-477-504A-65

; Sequence 65, Application US/08477504A

; Patent No. 5972353

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

; CITY: Tiburon

; STATE: California

; COUNTRY: USA

; ZIP: 94920

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/477,504A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/260,190

; FILING DATE: 15-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lauder, Leona L.

; REGISTRATION NUMBER: 30,863

; REFERENCE/DOCKET NUMBER: D-0021.3D

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-435-2034

; TELEFAX: 415-435-0727

; INFORMATION FOR SEQ ID NO: 65:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 105 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-477-504A-65

Query Match 0.3%; Score 80; DB 4; Length 105;

Best Local Similarity 85.6%; Pred. No. 9.5e-10;

Matches 89; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 10857 TTTTGTATTTTAGTAGACAGGTTTCACCGTGTGGCCAGGATGTTCTCAATCTCC 10916

Db 2 TTTTACATCTTTAGTAGACAGGTTTCACCATATTGGCCAGGCTGCTCTCAAACTCC 61

QY 10917 TTACCTCGTGATCCGCCCTCGCTGCTGCCAAAGTGTCTGGAT 10960

Db 62 TGACCTGTGTGATCCACGAGCTCGGCTCCCAAAAGTGTGGAT 105

## RESULT 3

US-08-486-756A-65

; Sequence 65, Application US/08486756A

; Patent No. 5981711

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

; CITY: Tiburon

; STATE: California

; COUNTRY: USA

; ZIP: 94920

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/486,756A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/260,190

; FILING DATE: 15-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lauder, Leona L.

; REGISTRATION NUMBER: 30,863

; REFERENCE/DOCKET NUMBER: D-0021.3C

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-435-2034

; TELEFAX: 415-435-0727

; INFORMATION FOR SEQ ID NO: 65:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 105 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-486-756A-65

Query Match 0.3%; Score 80; DB 4; Length 105;

Best Local Similarity 85.6%; Pred. No. 9.5e-10;

Matches 89; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 10857 TTTTGTATTTTAGTAGACAGGTTTCACCGTGTGGCCAGGATGTTCTCAATCTCC 10916

Db 2 TTTTACATCTTTAGTAGACAGGTTTCACCATATTGGCCAGGCTGCTCTCAAACTCC 61

QY 10917 TTACCTCGTGATCCGCCCTCGCTGCTGCCAAAGTGTCTGGAT 10960

Db 62 TGACCTGTGTGATCCACGAGCTCGGCTCCCAAAAGTGTGGAT 105

## RESULT 4

US-08-485-862B-65

; Sequence 65, Application US/08485862B

; Patent No. 5989838

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court





MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/481,658B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-481-658B-65

Query Match 0.3%; Score 76.2; DB 4; Length 105;  
Best Local Similarity 82.9%; Pred. No. 7.3e-09;  
Matches 87; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 9319 ATCCGAGCAGCTTTGGAGGCGGCGGATCAGGAGTCAGGAGATGGAGACCATC 9378  
|||||  
DB 105 ATCCGAGCAGCTTTGGAGGCGGCGGATCAGGAGTCAGGAGATGGAGACCATC 9378  
|||||

QY 9379 CTGCTTAACACGATGAACCCGCTCTCTACTAAATAACAAATA 9423  
|||||  
DB 45 CTGGCCAATATGTTGAACCCCTCTCTACTAAAGATGTAAAAA 1

RESULT 7  
US-08-477-504A-65/c  
Sequence 65, Application US/08477504A  
Patent No. 5972353  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,504A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:

NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-477-504A-65

Query Match 0.3%; Score 76.2; DB 4; Length 105;  
Best Local Similarity 82.9%; Pred. No. 7.3e-09;  
Matches 87; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 9319 ATCCGAGCAGCTTTGGAGGCGGCGGATCAGGAGTCAGGAGATGGAGACCATC 9378  
|||||  
DB 105 ATCCGAGCAGCTTTGGAGGCGGCGGATCAGGAGTCAGGAGATGGAGACCATC 9378  
|||||

QY 9379 CTGCTTAACACGATGAACCCGCTCTCTACTAAATAACAAATA 9423  
|||||  
DB 45 CTGGCCAATATGTTGAACCCCTCTCTACTAAAGATGTAAAAA 1

RESULT 8  
US-08-486-756A-65/c  
Sequence 65, Application US/08486756A  
Patent No. 5981711  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,756A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-0727  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-486-756A-65

Query Match 0.3%; Score 76.2; DB 4; Length 105;  
Best Local Similarity 82.9%; Pred. No. 7.3e-09;  
Matches 87; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 9319 ATCCGACACTTTGGGAGCGCGGCGGATCAGGATCAGGATGAGGACCATC 9378  
Db 105 ATCCGACACTTTGGGAGCGCGGCGGATCAGGATGAGGACCATC 9378  
QY 9379 CTGCTTAACAGATGAACCCCGTCTCTACTATAAAATACAAATA 9423  
Db 45 CTGCCCAATATGGTGAACCCCTGTCTCTACTATAAGATGTAATAAAA 1

## RESULT 9

US-08-485-862B-65/c  
Sequence 65, Application US/08485862B  
Patent No. 5989838  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,862B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727

INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-485-862B-65

Query Match 0.3%; Score 76.2; DB 4; Length 105;  
Best Local Similarity 82.9%; Pred. No. 7.3e-09;  
Matches 87; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 9319 ATCCGACACTTTGGGAGCGCGGCGGATCAGGATGAGGACCATC 9378  
Db 105 ATCCGACACTTTGGGAGCGCGGCGGATCAGGATGAGGACCATC 9378  
QY 9379 CTGCTTAACAGATGAACCCCGTCTCTACTATAAAATACAAATA 9423  
Db 45 CTGCCCAATATGGTGAACCCCTGTCTCTACTATAAGATGTAATAAAA 1

## RESULT 10

US-08-787-739-65/c  
Sequence 65, Application US/08787739  
Patent No. 6027887  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 96  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 369 Pine Street, Suite 610  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/787,739  
FILING DATE: 24-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,049  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/486,756  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/481,658  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,863  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/487,077  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-981-2034  
TELEFAX: 415-981-0332

INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-787-739-65

Query Match 0.3%; Score 76.2; DB 5; Length 105;  
Best Local Similarity 82.9%; Pred. No. 7.3e-09;  
Matches 87; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 9319 ATCCGAGCACTTTGGAGCGCGGCGGATCAGAGGTGAGGATGAGGACATC 9378  
|||||  
Db 105 ATCCGAGCACTTTGGAGCGCGGCGGATCAGAGGTGAGGATGAGGACATC 46  
|||||

QY 9379 CTGCTTAACAGATGAACCCCTCTCTACTAAATACAAATA 9423  
||| |||  
Db 45 CTGGCCAATATGTTGAACCCCTCTCTACTAAAGATGTAATAA 1

RESULT 11  
US-08-454-557C-91/c  
; Sequence 91, Application US/08454557C  
; Patent No. 5830670  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, Steven R.  
; REGISTRATION NUMBER: 36,203  
; REFERENCE/DOCKET NUMBER: 0609.3840003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 91:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 84 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
US-08-454-557C-91

Query Match 0.2%; Score 67.6; DB 3; Length 84;  
Best Local Similarity 89.0%; Pred. No. 6.5e-07;  
Matches 73; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 9309 CACGCTGTAAATCCAGCACTTTGGAGCGCGGATCAGAGGTGAGGATGAGGAT 9368  
|||||  
Db 83 CACGCTGTAAATCCAGCACTTTGGAGCGCGGATCAGAGGTGAGGATGAGGAT 24  
|||||

QY 9369 GGAGACCATCTCTGCTTAACAG 9390  
||| |||  
Db 23 CGACACCACTGATGACATG 2

RESULT 12  
US-08-340-426D-91/c  
; Sequence 91, Application US/08340426D  
; Patent No. 5948634

; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE: 14-NOV-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, Steven R.  
; REGISTRATION NUMBER: 36,203  
; REFERENCE/DOCKET NUMBER: 0609.3840002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 91:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 84 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
US-08-340-426D-91

Query Match 0.2%; Score 67.6; DB 4; Length 84;  
Best Local Similarity 89.0%; Pred. No. 6.5e-07;  
Matches 73; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 9309 CACGCTGTAAATCCAGCACTTTGGAGCGCGGATCAGAGGTGAGGATGAGGAT 9368  
|||||  
Db 83 CACGCTGTAAATCCAGCACTTTGGAGCGCGGATCAGAGGTGAGGATGAGGAT 24  
|||||

QY 9369 GGAGACCATCTCTGCTTAACAG 9390  
||| |||  
Db 23 CGACACCACTGATGACATG 2

RESULT 13  
US-08-450-673C-91/c  
; Sequence 91, Application US/08450673C  
; Patent No. 5948888  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25



Search completed: June 19, 2000, 03:08:43  
Job time: 396853 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2000, 03:02:32 ; Search time 17971 Seconds  
(without alignments)  
-1569.860 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_112000\_141000  
Perfect score: 29001  
Sequence: 1 TGTACTCTGGCTACCTCTG.....CTATAAAGTGAACAACCC 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 370290

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_bal.\*
- 2: gb\_ba2.\*
- 3: gb\_on.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pl1.\*
- 8: gb\_pl2.\*
- 9: gb\_pr1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_vi.\*
- 17: em\_fun.\*
- 18: em\_hum1.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_om.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pl.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_vi.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_ba1.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*
- 44: gb\_htg6.\*

- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*
- 50: gb\_pl3.\*
- 51: gb\_pr5.\*
- 52: gb\_btg8.\*
- 53: gb\_btg9.\*
- 54: gb\_btg10.\*
- 55: gb\_btg11.\*
- 56: gb\_btg12.\*
- 57: gb\_btg13.\*
- 58: gb\_btg14.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	99	0.3	108	10	HSLDLRN2
c 2	90.2	0.3	107	9	HUMALCE162
c 3	89.4	0.3	108	10	HSLDLRD1
c 4	89.4	0.3	108	10	HSLDLRD2
c 5	86.8	0.3	108	10	HSLDLRN2
c 6	85.4	0.3	103	9	HUMALCE221
c 7	83.2	0.3	108	10	HSLDLI12
c 8	82.6	0.3	108	11	HSU67803
c 9	81.2	0.3	107	9	HUMALCE162
c 10	79.8	0.3	103	9	HUMALCE221
c 11	79.4	0.3	108	10	HSLDLRD1
c 12	79.4	0.3	108	10	HSLDLRD2
c 13	76	0.3	103	13	HS8IC8R
c 14	75.6	0.3	110	11	HSU67807
c 15	75.2	0.3	103	13	HS8IC8R
c 16	74.4	0.3	104	9	HUMALCE272
c 17	74.2	0.3	108	9	HUMDLDO3M5
c 18	74.4	0.3	108	11	HSU67808
c 19	73.6	0.3	97	9	HUMLDLRA2
c 20	73.4	0.3	110	11	HSU67807
c 21	71.4	0.2	108	11	HSU67804
c 22	70.8	0.2	99	13	HUMUT7692A
c 23	70.8	0.2	108	10	HSLDLI12
c 24	70	0.2	95	13	HUMUT8002B
c 25	70.2	0.2	108	13	G32614
c 26	69.8	0.2	101	10	S79560
c 27	69.6	0.2	107	11	HSU67806
c 28	69.4	0.2	108	13	G43535
c 29	69	0.2	90	9	HUMLDLRFL
c 30	69.2	0.2	91	13	HUMUT8164A
c 31	68.4	0.2	79	10	S73203
c 32	68.2	0.2	101	10	S79560
c 33	67.8	0.2	95	10	HSTHPK1B
c 34	68	0.2	108	13	G43535
c 35	67.2	0.2	80	9	HUMBRFAE
c 36	66.6	0.2	91	13	HUMUT8164A
c 37	66.8	0.2	94	9	HUMHGAL
c 38	66.2	0.2	100	9	HUMGALNSA
c 39	66.2	0.2	100	9	HUMGALNSA
c 40	65.2	0.2	108	9	HUMDLDO3M5
c 41	64.4	0.2	90	10	HSU19407
c 42	64.6	0.2	100	10	HSLAS27
c 43	64.4	0.2	110	9	HUMALCE43
c 44	64	0.2	90	9	HUMLDLRFL
c 45	64	0.2	97	9	HUMLDLRA2

ALIGNMENTS

```

RESULT 1
HSLDLRN2 108 bp DNA PRI 20-MAY-1992
LOCUS Human LDL-receptor gene intron 14 fragment (normal gene).
DEFINITION X05250
ACCESSION X05250.1 GI:34337
VERSION Alu repetitive sequence; low density lipoprotein receptor.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Hovinga,J.R.,
Williamson,R. and Humphries,S.
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT See X05252 for deletion junction
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES
source Location/Qualifiers
1..108
/organism="Homo sapiens"
/db_xref="taxon:9606"
intron 1..108
/notes="intron XIV fragment"
BASE COUNT 28 a 23 c 39 g 18 t
ORIGIN

Query Match 0.3%; Score 99; DB 10; Length 108;
Best Local Similarity 95.3%; Pred. No. 1e-07; Indels 0; Gaps 0;
Matches 102; Conservative 0; Mismatches 5;

Qy 27523 ACAAATACAGGCGGTGGTGGCTGTCTGTATCCAGCTACTCAGGAGGCTGAG 27582
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 ACAAATATAGCAGGCGGTGGTGGCTGTCTGTATCCAGCTACTCAGGAGGCTGAG 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Qy 27583 GCAGAGAAATGCTTGACCCAGGAGCGGGAGGTGGCAGTGAGCCGA 27629
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 61 GCAGAGAAATGCTTGACCCAGGAGCGAGAGGTGGCAGTGAGCCGA 107
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 2
HUMALCEL62/c 107 bp ss-RNA PRI 15-APR-1994
LOCUS Human carcinoma cell-derived Alu RNA transcript, clone CE162.
DEFINITION M87924
ACCESSION M87924.1 GI:174871
VERSION Alu repeat.
KEYWORDS Alu repeat.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
REFERENCE 1 (bases 1 to 107)
AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
post-transcriptional selection of master sequences
JOURNAL J. Mol. Biol. (1992) In press
FEATURES
source Location/Qualifiers
1..107
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Nrera2D1"
/dev_stage="embryo"
/sex="male"
/tissue_type="carcinoma"
BASE COUNT 28 a 30 c 35 g 14 t
ORIGIN

```

```

Query Match 0.3%; Score 90.2; DB 9; Length 107;
Best Local Similarity 92.2%; Pred. No. 3.3e-06;
Matches 95; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 22252 TTTTCTGAGCAGGAGTCTTGTCTGTGCGCCAGGCTGGAATGGCAGATCTCGGC 22311
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 107 TTTTCTGAGCAGGAGTCTGCTGTGCGCCAGGCTGGAATGGCAGATCTCGGC 48
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Qy 22312 TCAGTCAACTCGGCTCCGGATTACGCCATTCCTCTCCGCC 22354
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 47 TCAGTCAACTCGGCTCCGGATTACGCCATTCCTCTCGGCC 5
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 3
HSLDLR1/c 108 bp DNA PRI 20-MAY-1992
LOCUS Human LDL-receptor mutated gene with intron 12 deletion junction.
DEFINITION X05249
ACCESSION X05249.1 GI:34335
VERSION Alu repetitive sequence; low density lipoprotein receptor.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Hovinga,J.R.,
Williamson,R. and Humphries,S.
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT *source: hypercholesterol aemia
See X05248 for corresponding normal gene sequence
In the defective LDL-receptor gene the deletion occurred between two
alu-repetitive sequences, that are in the same direction, the
deletion eliminates exons 13 and 14 and changes the reading frame
of the resulting spliced mRNA.
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES
source Location/Qualifiers
1..108
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="blood leukocytes from a patient with familial"
misc_feature 1..108
/notes="deletion junction region intron 12/ intron 15"
BASE COUNT 20 a 40 c 20 g 28 t
ORIGIN

Query Match 0.3%; Score 89.4; DB 10; Length 108;
Best Local Similarity 89.7%; Pred. No. 4.6e-06;
Matches 96; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 27523 ACAAATATAGCAGGCGGTGGTGGCTGTCTGTATCCAGCTACTCAGGAGGCTGAG 27582
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 108 ACAAATATAGCAGGCGGTGGTGGCTGTCTGTATCCAGCTACTCAGGAGGCTGAG 49
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Qy 27583 GCAGAGAAATGCTTGACCCAGGAGCGGGAGGTGGCAGTGAGCCGA 27629
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 48 GCAGAGAAATGCTTGACCCAGGAGCGAGAGGTGGTGGTGAGGCCGA 2
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 4
HSLDLR2 108 bp DNA PRI 20-MAY-1992
LOCUS Human LDL-receptor mutated gene with intron 14 deletion junction.
DEFINITION X05251
ACCESSION X05251.1 GI:34336
VERSION Alu repetitive sequence; low density lipoprotein receptor.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

```







Best Local Similarity 87.9%; Pred. No. 0.00021;  
Matches 87; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 22286 CTGGAATCAGTGGACAAATCTCGGCTCAGTGCACACCTCCGCTCCCGGATTCAGCCCAT 22345  
Db 103 CTGGAGTGAATGGACGACATCTCGGCTCAGTGCACACCTCCGCTCCCGGATTCAGCCCAT 44  
Qy 22346 TCTCTGCTCAACCTCCGAGTAGCTGGGACCAAGGC 22384  
Db 43 TCTCTGCTTAGCTTCCGCTGAGCTGGGATACAGGC 5

RESULT 11  
HSLDLRD1  
LOCUS Human LDL-receptor mutated gene with intron 12 deletion junction.  
DEFINITION X05249 PRI 20-MAY-1992  
ACCESSION X05249  
VERSION X05249.1 GI:34335  
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 108)  
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,  
Williamson,R. and Humphries,S.  
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in  
the low-density-lipoprotein-receptor gene. A possible mechanism for  
the defect in a patient with familial hypercholesterolaemia  
Eur. J. Biochem. 164 (1), 77-81 (1987)  
JOURNAL 87161901  
MEDLINE  
COMMENT \*source: hypercholesterol aemia  
See X05248 for corresponding normal gene sequence  
In the defective LDL-receptor gene the deletion occurred between two  
alu-repetitive sequences, that are in the same direction, the  
deletion eliminates exons 13 and 14 and changes the reading frame  
of the resulting spliced mRNA.  
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES  
source  
1..108  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="blood leukocytes from a patient with familial"  
intron  
1..108  
/note="intron XIV fragment"  
BASE COUNT 28 a 20 c 20 g 20 t  
ORIGIN

Query Match 0.3%; Score 79.4; DB 10; Length 108;  
Best Local Similarity 84.8%; Pred. No. 0.00025;  
Matches 89; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 22307 TCGGCTCACTGCAACCTCCGCTCCCGGATTCAGCCCATTCCTGCCTCAACCTCCCGA 22366  
Db 107 TCGCCTCACCACCACTCTGCTCTCTGGTTCAACCATTTTCTGCTCAGCCTCCCGA 48  
Qy 22367 GTAGTGGGACCAAGCGCGCCGCCACCAAGCCCAAGCTAAATTTT 22411  
Db 47 GTAGTGGGATTACAGGACCTGCCACCAAGCCTGGCTAAATTTT 3

RESULT 13  
HSLDLRD2/c  
LOCUS Human LDL-receptor mutated gene with intron 14 deletion junction.  
DEFINITION X05251 PRI 20-MAY-1992  
ACCESSION X05251  
VERSION X05251.1 GI:34336  
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 108)  
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,  
Williamson,R. and Humphries,S.  
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in  
the low-density-lipoprotein-receptor gene. A possible mechanism for  
the defect in a patient with familial hypercholesterolaemia  
Eur. J. Biochem. 164 (1), 77-81 (1987)  
JOURNAL 87161901  
MEDLINE  
COMMENT \*source: hypercholesterol aemia  
See X05248 for corresponding normal gene sequence  
In the defective LDL-receptor gene the deletion occurred between two  
alu-repetitive sequences, that are in the same direction, the  
deletion eliminates exons 13 and 14 and changes the reading frame  
of the resulting spliced mRNA.  
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES  
source  
1..108  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="blood leukocytes from a patient with familial"  
misc\_feature  
1..108  
/note="deletion junction region intron 12/ intron 15"  
BASE COUNT 20 a 40 c 20 g 28 t  
ORIGIN

Query Match 0.3%; Score 79.4; DB 10; Length 108;  
Best Local Similarity 84.8%; Pred. No. 0.00025;  
Matches 89; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 22307 TCGGCTCACTGCAACCTCCGCTCCCGGATTCAGCCCATTCCTGCCTCAACCTCCCGA 22366  
Db 2 TCGCCTCACCACCACTCTGCTCTCTGGTTCAACCATTTTCTGCTCAGCCTCCCGA 61  
Qy 22367 GTAGTGGGACCAAGCGCGCCGCCACCAAGCCCAAGCTAAATTTT 22411  
Db 62 GTAGTGGGATTACAGGACCTGCCACCAAGCCTGGCTAAATTTT 106

RESULT 12  
HSLDLRD2/c  
LOCUS Human LDL-receptor mutated gene with intron 14 deletion junction.  
DEFINITION X05251 PRI 20-MAY-1992  
ACCESSION X05251  
VERSION X05251.1 GI:34336  
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 108)  
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,  
Williamson,R. and Humphries,S.  
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in  
the low-density-lipoprotein-receptor gene. A possible mechanism for  
the defect in a patient with familial hypercholesterolaemia  
Eur. J. Biochem. 164 (1), 77-81 (1987)  
JOURNAL 87161901  
MEDLINE  
COMMENT \*source: hypercholesterol aemia  
See X05250 for corresponding normal gene sequence  
In the defective LDL-receptor gene the deletion occurred between two  
alu-repetitive sequences, that are in the same direction, the  
deletion eliminates exons 13 and 14 and changes the reading frame  
of the resulting spliced mRNA.  
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES  
Location/Qualifiers  
1..108  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="blood leukocytes from a patient with familial"  
intron  
1..108  
/note="intron XIV fragment"  
BASE COUNT 28 a 20 c 40 g 20 t  
ORIGIN

Query Match 0.3%; Score 79.4; DB 10; Length 108;  
Best Local Similarity 84.8%; Pred. No. 0.00025;  
Matches 89; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 22307 TCGGCTCACTGCAACCTCCGCTCCCGGATTCAGCCCATTCCTGCCTCAACCTCCCGA 22366  
Db 107 TCGCCTCACCACCACTCTGCTCTCTGGTTCAACCATTTTCTGCTCAGCCTCCCGA 48  
Qy 22367 GTAGTGGGACCAAGCGCGCCGCCACCAAGCCCAAGCTAAATTTT 22411  
Db 47 GTAGTGGGATTACAGGACCTGCCACCAAGCCTGGCTAAATTTT 3

RESULT 13  
HSLDLRD2/c  
LOCUS Human sequence tagged site 8IC8R DNA from 19q13.  
DEFINITION X57789 STS  
ACCESSION X57789  
VERSION X57789.1 GI:23938  
KEYWORDS STS; myotonic dystrophy.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 103)  
AUTHORS Aldridge,F.L.  
TITLE Direct Submission  
JOURNAL Submitted (12-FEB-1991) F.L. Aldridge, ICI Pharmaceuticals,  
Alderley Park, Macclesfield, Cheshire, SK10 4TG, UK  
REFERENCE 2 (bases 1 to 103)  
AUTHORS Butler,R., Riley,J.H., Ogilvie,D.J., Anand,R., Buxton,J.,  
Davies,J., Johnson,K. and Markham,A.F.  
TITLE Two sequence-tagged sites defining the ends of a 380 kb YAC clone  
from 19q13  
JOURNAL Nucleic Acids Res. 19 (17), 4787 (1991)  
MEDLINE 91367697  
COMMENT See also X57788 for STS 8IC8L.  
FEATURES Location/Qualifiers  
source  
1..103  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19q13"  
/germline  
/clone\_lib="YAC library: ICI"  
/clone="8IC8"  
BASE COUNT 23 a 28 c 23 g 22 t 1 others  
ORIGIN









```
RESULT 2
X12095
ID X12095 standard; DNA; 108 BP.
AC X12095;
DE 30-MAR-1999 (first entry)
DE Human biallelic polymorphic DNA fragment TIGR-A003M18a.
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW treatment; marker; ss.
OS Homo sapiens.
PN W09820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; US-030455.
PR 06-NOV-1997; UO-030455.
PA (WHEED ) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
DR WPI; 98-286974/25.
DR New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
PS Claim 1; Page 219; 31Opp; English.
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC X09121-X10268. The base occupying the polymorphic site is indicated by
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberosus sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC segments can also be used to produce medicaments for the treatment or
CC prophylaxis of such diseases.
SQ Sequence 108 BP; 19 A; 23 C; 28 G; 37 T;

Query Match 0.2%; Score 68.6; DB 1; Length 108;
Best Local Similarity 82.8%; Pred. No. 0.013;
Matches 77; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

QY 12486 TGTATTTTGTAGAGAGCGGGTTTACCAATGTTGGCAGGCTGTCTCAAACTCCTGAC 12545
Db 1 TGTCTTTTGTAGAGATGAGGTTTCTCTGTGTTGGCAGGATGCTCGAACTCCTGAC 60

QY 12546 CTCAGTGATCCACTGCCTCGGCTCCCAAAA 12578
Db 61 TTCAAGTGATCCGTCTGCTTGGCTCCCAAAA 93

RESULT 3
T25009
ID T25009 standard; cDNA to mRNA; 108 BP.
AC T25009;
DE 07-NOV-1996 (first entry)
DE Human gene signature HUMGS07131.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATSU/) MATSUBARA K.
PA (OKUBU/) OKUBO K.

RESULT 4
T24892
ID T24892 standard; cDNA to mRNA; 100 BP.
AC T24892;
DE 05-NOV-1996 (first entry)
DE Human gene signature HUMGS06998.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATSU/) MATSUBARA K.
PA (OKUBU/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
DR Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1720; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 108 BP; 34 A; 31 C; 26 G; 15 T;
```

```
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
DR Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1748; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 108 BP; 34 A; 31 C; 26 G; 15 T;

Query Match 0.2%; Score 63.4; DB 1; Length 108;
Best Local Similarity 73.8%; Pred. No. 0.082;
Matches 79; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 27591 ATTGCTTTGAACCCAGGAGCGGGTTCAGTGTGAGCGGAATCGCCACTGCTCCAG 27650
Db 2 ATCGCTTGAGCCATGAGGCGCAAGGTCGAGTGCATGTCACGCCACTGNATTCCAG 61

QY 27651 CCTGGGTAAACAGACGAGGCTCTGTTTCAAAAATAAATAATACATA 27697
Db 62 CCTGAGTGACAGAGCAAGACCTGTTGAAAACACACACACANCAA 108

RESULT 4
T24892
ID T24892 standard; cDNA to mRNA; 100 BP.
AC T24892;
DE 05-NOV-1996 (first entry)
DE Human gene signature HUMGS06998.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATSU/) MATSUBARA K.
PA (OKUBU/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
DR Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1720; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
```





PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues

PS Claim 1; Page 1748; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in T19001-T26837 and which is able to hybridize to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost

CC all the 3'-oriented cDNAs hybridize with specific mRNAs. Each library

CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be

CC determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.

CC Sequence 108 BP; 34 A; 26 G; 15 T;

SQ

[illegible]

RESULT	8	
T21566		
ID	T21566 standard; cDNA to mRNA; 87 BP.	
AC	T21566;	
DC	03-AUG-1996 (first entry)	
DE	Human gene signature HUMG502944.	
KE	Gene signature; messenger RNA; mRNA; relative abundance; frequency;	
KW	human; cloning; mapping; non-biased library; diagnosis; detection;	
KW	cell typing; abnormal cell function; ss.	
OS	Homo sapiens.	
PN	WO951472-A1.	
PD	01-JUN-1995.	
PF	11-NOV-1994; J01916.	
PR	12-NOV-1993; JP-355504.	
PA	(MATS/) MATSUBARA K.	
PK	(OKUB/) OKUBO K.	
PI	Matsubara K. Okubo K;	
DR	WPI; 95-206931/27.	
PT	Identifying gene signatures in 3'-directed human cDNA library - e.g.	
PT	for diagnosis of abnormal cell function, by preparing cDNA that	
PT	reflects relative abundance of corresp. mRNA in specific human	
PT	tissues	

Claim 1; Page 914; 2245pp; Japanese.  
A single-stranded DNA (or its complementary strand or the corresp.  
double-stranded DNA) which comprises one of the 7837 "GS" sequences  
given in T19001-T76837 and which is able to hybridize to part of  
human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
sequences were obtained from 3'-directed cDNA libraries prepared  
from various human tissues; synthesis of cDNA was initiated from the  
3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
untranslated sequence is unique to a particular mRNA species, almost  
all the 3'-oriented cDNAs hybridize with specific mRNAs. Each library  
is constructed so as to reflect accurately the relative abundance of  
different mRNAs in the particular tissue from which it was derived.  
The appearance frequency of a given GS in a cDNA library can be  
determined (esp. using primers and probes derived from the GS  
sequences) as a means of diagnosing abnormal cell function or for  
recognising different cell types

```

SQ      Sequence      87 BP;  35 A;   21 C;   16 G;   13 T;
          Query Match      0.2%;   Score 58;   DB 1;   Length 87;
          Best Local Similarity 79.8%;   Pred. No. 0.55;
          Matches 67;   Conservative 0;   Mismatches 17;   Indels
QY 23100  GATCACTTGGATCCAGAGTTTGAGACAGAGCTGGTCAACATGCGGAACCT
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1  GATCGCTTGANCCAGAGGTTTAAACCGCCGCGGAGACATGGCGGAACCC
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 23160  CAAAAATAAAAAAATTTGTCAGG 23183
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      61  CAAAAAATACAGAAATNAGCCAAG 84

RESULT 9
T21566/c
ID T21566 standard; cDNA to mRNA; 87 BP.
AC T21566:

```

RESULT	9	
T21566/c		
ID	T21566 standard; cDNA to mRNA; 87 BP.	
AC	T21566;	
DT	03-AUG-1996 (first entry)	
DE	Human gene signature HUMGS02944.	
KE	Gene signature; messenger RNA; mRNA; relative abundance; frequency;	
KW	human; cloning; mapping; non-biased library; diagnosis; detection;	
KW	cell typing; abnormal cell function; ss.	
OS	Homo sapiens.	
PN	WO9514772-A1.	
PD	01-JUN-1995.	
PF	11-NOV-1994; J01916.	
PR	12-NOV-1993; JP-355504.	
PA	(MATS/) MATSUBARA K.	
PA	(OKUB/) OKUBO K.	
PI	Matsubara K. Okubo K;	
DR	WPI; 95-206931/27.	
PT	Identifying gene signatures in 3'-directed human cDNA library - e.g.	
PT	for diagnosis of abnormal cell function, by preparing cDNA that	
PT	reflects relative abundance of corresp. mRNA in specific human	
PT	tissues	
PS	Claim 1; Page 914; 22425pp; Japanese.	

A single-stranded DNA (or its complementary strand or the corresp.  
double-stranded DNA) which comprises one of the 7837 "GS" sequences  
given in T19001-T26837 and which is able to hybridise to part of  
human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
sequences were obtained from 3'-directed cDNA libraries prepared  
from various human tissues; synthesis of cDNA was initiated from the  
3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
untranslated sequence is unique to a particular mRNA species, almost  
all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
is constructed so as to reflect accurately the relative abundance of  
different mRNAs in the particular tissue from which it was derived.  
The appearance frequency of a given GS in a cDNA library can be  
determined (esp. using primers and probes derived from the GS  
sequences) as a means of diagnosing abnormal cell function or for  
recognising different cell types.

	87 BP;	35 A;	21 C;	16 G;	13 T;
SQ					

[illegible]

RESULT 10  
T26828  
ID T26828 standard: cDNA to mRNA; 108 BP.

T26828;  
14-NOV-1996 (first entry)  
Human gene signature HUMGS09078.  
Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
human; Cloning; mapping; non-biased library; diagnosis; detection;  
cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN W09514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATSU) MATSUBARA K.  
PI (OKUBU) OKUBO K.  
DR WPI; 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1; Page 2182; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 108 BP; 18 A; 33 C; 23 G; 28 T;

Query Match 0.2%; Score 58.2; DB 1; Length 108;  
Best Local Similarity 77.5%; Pred. No. 0.52;  
Matches 69; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 22461 GATCTCTGACCTCATGATCTGCCACCTCAGCCCTCCCAAGTGTAGGATCAGCGCAT 22520  
|||||  
1 GATCTCTGACCTCATGATCTGCCCGCGTGTGCGCTCCCATGATGTGCGGTTACAGGCAT 60  
|||||

Qy 22521 GAGCCACCGCGCCGCTACTGACTTTT 22549  
|||||  
61 GAGCCACCGCGCGGCTGTTTATTCT 89  
|||||

RESULT 11  
X12087/c  
ID X12087 standard; DNA; 100 BP.  
AC X12087;  
DT 30-MAR-1999 (first entry)  
DE Human biallelic polymorphic DNA fragment EST98276a.  
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;  
KW detection; phenotypic typing; characteristic; infection; hereditary;  
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;  
KW treatment; marker; ss.  
OS Homo sapiens.  
PN W09820165-A2.  
PD 14-MAY-1998.  
PF 05-NOV-1997; U20313.  
PR 06-NOV-1996; US-030455.  
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.  
PI Hudson T, Lander ES, Wang D;  
DR WPI; 98-286974/25.  
PT New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease  
PS Claim 1; Page 219; 310pp; English.

CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
CC markers which have been isolated using the primers represented in  
CC X09121-X10268. The base occupying the polymorphic site is indicated by  
CC the appropriate IUPAC-TUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.  
SQ Sequence 100 BP; 21 A; 25 C; 22 G; 31 T;

Query Match 0.2%; Score 57; DB 1; Length 100;  
Best Local Similarity 72.7%; Pred. No. 0.8;  
Matches 72; Conservative 1; Mismatches 26; Indels 0; Gaps 0;

Qy 27408 GTGGCTCACCTGTAATCCAGCACCTTGGGAGCCCAAGGTAAGCAGATCATTGAGGT 27467  
|||||  
99 GTGGCTCACCTGTAATCCCTGGCACCTTAGGAGGCTTAGGAAGGAGGATCTTTGAAC 40  
|||||

Qy 27468 CAGGAGTTAGAGCAGCAGCTCTGGCCCAACATAGTGAAATC 27506  
|||||  
39 CAGGAGCTCAGACCAKCTGGGAACATAGCAAGCATC 1  
|||||

RESULT 12  
X12085/c  
ID X12085 standard; DNA; 100 BP.  
AC X12085;  
DT 30-MAR-1999 (first entry)  
DE Human biallelic polymorphic DNA fragment EST98276c.  
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;  
KW detection; phenotypic typing; characteristic; infection; hereditary;  
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;  
KW treatment; marker; ss.  
OS Homo sapiens.  
PN W09820165-A2.  
PD 14-MAY-1998.  
PF 05-NOV-1997; U20313.  
PR 06-NOV-1996; US-030455.  
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.  
PI Hudson T, Lander ES, Wang D;  
DR WPI; 98-286974/25.  
PT New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease  
PS Claim 1; Page 218; 310pp; English.  
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
CC markers which have been isolated using the primers represented in  
CC X09121-X10268. The base occupying the polymorphic site is indicated by  
CC the appropriate IUPAC-TUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular



PI Hudson T, Lander ES, Wang D;  
DR WPI; 98-286974/25.  
PT New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease  
PS Claim 1; Page 219; 310pp; English.  
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
CC markers which have been isolated using the primers represented in  
CC X09121-X10268. The base occupying the polymorphic site is indicated by  
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tubercous sclerosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.  
SQ Sequence 100 BP; 21 A; 25 C; 22 G; 31 T;

Query Match 0.2%; Score 56.4; DB 1; Length 100;  
Best Local Similarity 75.0%; Pred. No. 0.99;  
Matches 69; Conservative 1; Mismatches 22; Indels 0; Gaps 0;  
QY 3246 GCATGTTGGCCAGGTGATCCGGAACCTCGTATTTCTGGTAATCGCCCGCTCAGCCT 3305  
Db 8 GCTATGTTTCCAGGATGCTCTTGAGCTCCTGGTTCAAAACATCCTCTTCTTACGCCT 67  
QY 3306 CTTAAAGTCTTGAATTACAGCGGTGAGTCAC 3337  
Db 68 CCTAAAGTCCAGGATTATAGGTGTGAGTCAC 99

Search completed: June 19, 2000, 11:47:50  
Job time: 427781 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: June 18, 2000, 22:02:37 ; Search time 8514.75 Seconds  
(without alignments)  
13805.165 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_112000\_141000  
Perfect score: 29001  
Sequence: 1 TGTACTCTGGCTACCTCTG.....CTATAAAGTGAACAACCC 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues  
Total number of hits satisfying chosen parameters: 156056

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : EST:  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result % Query

No.	Score	Match	Length	DB	ID	Description
1	93.2	0.3	106	37	AA703692	AA703692 ag81a10.r
2	91.4	0.3	109	30	AA243009	AA243009 zt25h02.s
3	90.4	0.3	106	30	AA250812	AA250812 z806a05.s
c 4	88.6	0.3	103	84	B48914	B48914 RPC111-4A12
c 5	88.2	0.3	101	39	AA835205	AA835205 ak64h01.s
c 6	88.2	0.3	110	30	AA244245	AA244245 nc07a04.s
c 7	87.8	0.3	101	33	AA381369	AA381369 EST94442
c 8	87.8	0.3	108	34	B65160	B65160 CIT-HSP-201
c 9	86.6	0.3	102	36	AA654562	AA654562 nt75f10.s
c 10	86.2	0.3	107	35	AA565533	AA565533 nk42b11.s
c 11	86.4	0.3	107	39	AA828124	AA828124 od71a07.s
c 12	86.6	0.3	110	94	AQ003188	AQ003188 RPC111-1D
c 13	86	0.3	102	36	AA654562	AA654562 nt75f10.s
c 14	86	0.3	103	108	AQ582186	AQ582186 RPC1-11-4
c 15	85.4	0.3	103	108	AQ535244	AQ535244 RPC1-11-3
c 16	85.2	0.3	109	94	AQ028426	AQ028426 CIT-HSP-2
c 17	85	0.3	110	30	AA244245	AA244245 nc07a04.s
c 18	85.2	0.3	110	39	AA897366	AA897366 am06h02.s
c 19	84.4	0.3	105	21	T94466	T94466 ve35b02.r1
c 20	84.6	0.3	107	103	AQ240182	AQ240182 CIT-HSP-2
c 21	84.2	0.3	106	108	AQ544957	AQ544957 CITBI-EI-
c 22	84.4	0.3	110	106	AQ386882	AQ386882 RPC111-13
c 23	83.8	0.3	103	94	AQ028649	AQ028649 CIT-HSP-2
c 24	83.8	0.3	103	108	AQ535244	AQ535244 RPC1-11-3
c 25	83.6	0.3	106	63	AI991750	AI991750 wt48e01.x
c 26	83.6	0.3	106	63	AI991750	AI991750 wt48e01.x
c 27	83.6	0.3	109	84	B17434	B17434 345K2.IVB C
c 28	83.2	0.3	96	92	AQ936334	AQ936334 RPC1-11-S
c 29	83.2	0.3	104	105	AQ321855	AQ321855 RPC111-11
c 30	83.4	0.3	109	84	B17434	B17434 345K2.TVB C
c 31	83	0.3	101	33	AA381369	AA381369 EST94442
c 32	82.8	0.3	103	38	AA807640	AA807640 nx08b05.s
c 33	82.2	0.3	103	108	AQ584425	AQ584425 RPC1-11-4
c 34	82.2	0.3	105	109	AQ637292	AQ637292 RPC1-11-4
c 35	82.4	0.3	108	84	B32951	B32951 HS-1016-A1-
c 36	81.6	0.3	104	105	AQ321855	AQ321855 RPC111-11
c 37	81.8	0.3	106	30	AA250812	AA250812 z806a05.s
c 38	82	0.3	106	38	AA812141	AA812141 eb48h02.s
c 39	82	0.3	106	50	AI700000	AI700000 tt36a10.x
c 40	81.6	0.3	105	30	AA218889	AA218889 zq15d04.s
c 41	81	0.3	105	28	AA078003	AA078003 7H12D08 C
c 42	81.2	0.3	110	106	AQ386882	AQ386882 RPC111-13
c 43	80.6	0.3	107	33	AA385808	AA385808 EST9495
c 44	80.2	0.3	101	94	AQ076649	AQ076649 CIT-HSP-2
c 45	80.2	0.3	101	105	AQ260734	AQ260734 CITBI-EI-

## ALIGNMENTS

```

RESULT 1
AA703692
LOCUS
DEFINITION ag81a10.r1 StrataGene hNT neuron (#937233) Homo sapiens cDNA clone
IMAGE:1140858 5' similar to contains Alu repetitive element; , mRNA
sequence.
ACCESSION AA703692
VERSION AA703692.1 GI:2713610
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le.N., Lennon,G., Marra,M.,
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397630.

```

```

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 53.
Location/Qualifiers
1. .106
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:1140858"
/clone_lib="Stratagene hNT neuron (#937233)"
/lab_stage="hNT neurons"
/dev_host="SOLR (kanamycin resistant)"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dT.
Differentially, post mitotic hNT neurons. Average insert
size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
GAATCGGCACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT 19 a 29 c 29 g 29 t
ORIGIN
Query Match 0.3%; Score 93.2; DB 37; Length 106;
Best Local Similarity 92.5%; Pred. No. 0.15;
Matches 98; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 22416 TTTTATTAGACAGAGTTTCACCGTGTAGCCGGATGCTCTCGATCTCCTGACCTCA 22475
Db 1 TTTTATTAGACAGAGTTTCACCGTGTAGCCGGATGCTCTCGATCTCCTGACCTCG 60
QY 22476 TGATCTGCCACCTCAGCCTCCCAAGTGTAGGATCAGAGCGCATG 22521
Db 61 TGATCTGCCCGCTCAGCCTCCCAAGTGTGGGATTACAGGCGTG 106
RESULT 2
AA243009 109 bp mRNA EST 11-MAR-1998
LOCUS AA243009
DEFINITION zt25h02.s1 StrataGene NT2 neuronal precursor 937230 Homo sapiens
cDNA clone IMAGE:664467 3' similar to contains Alu repetitive
element;contains element LTR1 repetitive element ; , mRNA sequence.
ACCESSION AA243009
VERSION AA243009.1 GI:1873869
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le.N., Lennon,G., Marra,M.,
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT On Dec 3, 1996 this sequence version replaced gi:1126869.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1127 Std Error: 0.00
Seq primer: -41m13 fwd. ET from Amersham
High quality sequence stop: 102.

```



FEATURES  
source

Location/Qualifiers  
1. .109  
/organism="Homo sapiens"  
/db\_xref="GDB:5426481"  
/db\_xref="taxon:9606"  
/clone="IMAGE:664467"  
/tissue\_type="neuroepithelial cells"  
/dev\_stage="Ntera-2 neuroepithelial cells"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: brain; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Uninduced, exponentially growing neuroepithelial  
cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG  
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3'"

BASE COUNT 19 a 30 c 30 g 30 t  
ORIGIN

Query Match 0.3%; Score 91.4; DB 30; Length 109;  
Best Local Similarity 89.9%; Pred. No. 0.24; Indels 0; Gaps 0;  
Matches 98; Conservative 0; Mismatches 11;

Qy 22413 GTATTTTAGTAGACAGGGTTTCACCGTGTAGCCGGGATGGTCTCGATCTCCTGACC 22472  
|||||  
Db 1 GTATTTTAGTAGACAGGGTTTCACCGTGTAGCCAGGATGGTCTGATCTCCTTACC 60  
|||||

Qy 22473 TCATGATCTGCCACCTCAGCGCTCCCAAGTGTAGGATACAGCGATG 22521  
|||||  
Db 61 TCGTGATCGGCCACCTCGCGCTCCCAAGTGTGGGATTACAGCGGTG 109  
|||||

RESULT 3  
AA250812

LOCUS AA250812 106 bp mRNA EST 15-AUG-1997  
DEFINITION zs06a05.s1 NCI-CGAP GCB1 Homo sapiens cDNA clone IMAGE:684368 3'  
similar to contains Alu repetitive element; contains element MER22  
repetitive element ;, mRNA sequence.

ACCESSION AA250812  
VERSION AA250812.1 GI:1885774  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 106)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced gi:1407356.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 537 Std Error: 0.00  
Seq primer: -41m13 fwd. ET from Amersham  
High quality sequence stop: 87.  
Location/Qualifiers  
1. .106  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:684368"  
/clone\_lib="NCI-CGAP GCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="Vector: p773D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),  
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman

FEATURES  
source

Location/Qualifiers  
1. .106  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:684368"  
/clone\_lib="NCI-CGAP GCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="Vector: p773D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),  
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman

BASE COUNT 20 a 28 c 31 g 27 t  
ORIGIN

Query Match 0.3%; Score 90.4; DB 30; Length 106;  
Best Local Similarity 94.0%; Pred. No. 0.31;  
Matches 94; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 12497 TAGAGACGGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCTTGACCTCAGGTGATC 12556  
|||||  
Db 2 TAGAGACGGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCTTGACCTCAGGTGATC 61  
|||||

Qy 12557 CACCTGCTCGCCCTCCCAAAATGCTGAGATTACAGGTGT 12596  
|||||  
Db 62 CACTTGCCTTGGCCTCCCAAAAGTGTGGGATTACAGGTGT 101  
|||||

## RESULT 4

LOCUS B48914/c B48914 103 bp DNA GSS 08-APR-1999  
DEFINITION RPC111-4A12.TP RPCI-11 Homo sapiens genomic clone RPCI-11-4A12,  
genomic survey sequence.

ACCESSION B48914  
VERSION B48914.1 GI:2601151  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 103)  
Adams, M.D., Rounsley, S.D., Field, C.E., Bass, S., Linher, K., and  
Golden, K., Berry, K., Granger, D., Sub, E., Wible, C., de Jong, P. and  
Venter, J.C.  
Use of BAC End Sequences for Sequence-Ready Map Building  
Unpublished (1997)  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdamads@tigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
([pieter@dejong.med.buffalo.edu](mailto:pieter@dejong.med.buffalo.edu)). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from  
Research Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: SP6  
Class: BAC ends.  
Location/Qualifiers  
1. .103  
/organism="Homo sapiens"  
/db\_xref="GDB:7501163"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-4A12"  
/clone\_lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPC111 Human Male BAC Library"

BASE COUNT 30 a 28 c 30 g 15 t  
ORIGIN



```

RESULT 7
AA381369/c
LOCUS
DEFINITION
  EST94442 Activated T-cells 1 Homo sapiens cDNA 5' end similar to
  EST containing Alu repeat, mRNA sequence.
ACCESSION
  AA381369
VERSION
  AA381369.1 GI:2033689
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 (bases 1 to 101)
  Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
  Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
  White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
  Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
  Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
  Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
  Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
  Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
  Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
  Small,K.V., Spriggs,T.A., Otterback,T.R., Weidman,J.F., Li,Y.,
  Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
  Dimke,D., Feng,D.-F., Fertie,A., Fischer,C., Hastings,G.A.,
  He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
  Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
  Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
  Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
  Fraser,C.M. and Venter,J.C.
  Initial assessment of human gene diversity and expression patterns
  based upon 83 million nucleotides of cDNA sequence
  Nature 377 (6547 Suppl), 3-174 (1995)
  12140200
  On Sep 12, 1996 this sequence version replaced gi:1407448.
  Other ESTs: THCI70052
  Contact: Kerlavage, AR
  Bioinformatics
  The Institute for Genomic Research
  9712 Medical Center Drive, Rockville, MD 20850 USA
  Tel: 3018699056
  Fax: 3018699423
  Email: arkerlavet@tigr.org
  For clone availability, additional sequence and expression
  information related to this EST, please check the TIGR Human Gene
  Index (http://www.tigr.org/tldb/hgi/hgi.html)
  Seq primer: M13 Reverse.
  Location/Qualifiers
    1..101
    /organism="Homo sapiens"
    /db_xref="ATCC (inhost):185728"
    /db_xref="taxon:9606"
    /clone_lib="Activated T-cells 1"
    /cell_type="T-lymphocyte"
    /dev_stage="adult"
    /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
    XhoI"
  BASE COUNT 18 a 36 c 20 g 25 t 2 others
  ORIGIN
    Query Match 0.3%; Score 87.8; DB 33; Length 101;
    Best Local Similarity 91.1%; Pred. No. 0.62;
    Matches 92; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

  QY 15362 AAAATTAGCTGGGTGGTGGCGGCACCTGTAATCCAGCTTAATCAGGAGCTGAGGCA 15421
  |||||
  Db 101 AAAATTAGCTGGAGTGGTGGCGGCGCCTGTAATCCAGCTTAATCAGGAGCTGAGGCA 42
  |||||

  QY 15422 GGAGATCGCTTGACCCAGGAGGAGGAGGTTCAGTGAGC 15462
  |||||
  Db 41 GGANAATTGCTTGAACCCAGGAGGAGGAGGTTCGAATGAGC 1

RESULT 8
B65160 108 bp DNA GSS 21-JUN-1998
LOCUS
DEFINITION
  CIT-HSP-2017G2.TFB CIT-HSP Homo sapiens genomic clone 2017G2,
  genomic survey sequence.
ACCESSION
  B65160
VERSION
  B65160.1 GI:2639138
KEYWORDS
  GSS.
SOURCE
  human.
ORGANISM
  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 (bases 1 to 108)
  Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
  Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
  Simon,M. and Venter,J.C.
  Use of a random BAC End Sequence Database for Sequence-Ready Map
  Building
  Unpublished (1997)
  Other GSSs: CIT-HSP-2017G2.TFB
  Contact: Mark Adams
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: mdadams@tigr.org
  Clones are available from Research Genetics (info@resgen.com). BAC
  end search page:
  http://www.tigr.org/tldb/hungen/bac\_end\_search/bac\_end\_search.html
  Seq primer: M13 Reverse
  Class: BAC ends.
  Location/Qualifiers
    1..108
    /organism="Homo sapiens"
    /db_xref="GDB:7043860"
    /clone="2017G2"
    /clone_lib="CIT-HSP"
    /sex="Male"
    /cell_type="Sperm"
    /note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
    HindIII"
  BASE COUNT 26 a 27 c 34 g 21 t
  ORIGIN
    Query Match 0.3%; Score 87.8; DB 84; Length 108;
    Best Local Similarity 88.8%; Pred. No. 0.6;
    Matches 95; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

  QY 27541 TGGTGGCATGTCCTGTATCCAGCTACTCAGGAGGCTGAGGCAAGAAATCTCTTGA 27600
  |||||
  Db 1 TGGTGGCATGTCCTGTATCCAGCTACTCAGGAGGCTGAGGCAAGAAATCTCTTGA 60
  |||||

  QY 27601 CCCAGAGCGCGGTGTCAGTGAGCCGGAATCGCCCACTGCACTC 27647
  |||||
  Db 61 CCCGGAGGTGGAGGTTCAGTGAGCCAGATCATACCACTGCACAC 107
  |||||

RESULT 9
AA654562/c
LOCUS
DEFINITION
  nt75f10.s1 NCI_CGAP_Pr3 Homo sapiens cDNA clone IMAGE:1204363
  similar to contains Alu repetitive element;contains element MER22
  repetitive element ;, mRNA sequence.
ACCESSION
  AA654562
VERSION
  AA654562.1 GI:2590716
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens

```

**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**  
**COMMENT**

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 102)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 On Sep 12, 1996 this sequence version replaced gi:1393451.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,  
 M.D., Michael Emmert-Buck, M.D., Ph.D.  
 CDNA Library Arrayed by: David B. Krizman, Ph.D.  
 DNA Sequencing by: Genome Systems Inc., Greg Lennon, Ph.D.  
 Clone distribution: Washington University Genome Sequencing Center  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

**FEATURES**  
 source  
 Location/Qualifiers  
 1..102  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1204363"  
 /clone\_lib="NCI-CGAP\_Pr3"  
 /sex="Male"  
 /dev\_stage="45 years old"  
 /lab\_host="Dhl08"  
 /note="vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st strand cDNA was primed with oligo(dT)17 on 50 ng of DNase-treated, total cellular RNA obtained from 5,000-10,000 microdissected cells histologically determined to be fully malignant prostate cancer cells. Double-stranded cDNA was ligated to EcoRI adaptors, 5 cycles of PCR applied to the cDNA with an adaptor-specific primer, and the resulting PCR product subcloned into pAMP10 by the UDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman."  
 22 a 32 c 27 g 21 t

Query Match 0.3%; Score 86.6; DB 36; Length 102;  
 Best Local Similarity 91.1%; Pred. No. 0.83;  
 Matches 92; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 15247 GCTCATGCTCTAATACAGCAGCTTTGGAGCGCGATGTGGTGATCACCTGAGTCAG 15306  
 Db 101 GCTACTCTCTATATCCAGCAGCTTTGGAGCGCGATGTGGTGATCACCTGAGTCAG 42

QY 15307 GAGTTTGACGACAGCTGGCCAAACATGGTGAACCTCATCT 15347  
 Db 41 GAGTTTGACGACAGCTGGCCAAACATGGTGAACCTCATCT 1

**RESULT** 10  
**AA565533/c**  
**LOCUS**  
**DEFINITION**  
**ACCSSION**  
**VERSION**  
**KEYWORDS**  
**SOURCE**  
**ORGANISM**  
**REFERENCE**  
**AUTHORS**

AA565533 107 bp mRNA EST 08-SEP-1997  
 nk42b11.s1 NCI-CGAP\_GC2 Homo sapiens cDNA clone IMAGE:1016157 3'  
 similar to contains Alu repetitive element;; mRNA sequence.  
 AA565533  
 AA565533.1 GI:2337172  
 EST.  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 107)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 On Jan 17, 1998 this sequence version replaced gi:1899815.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher A. Moskaluk, M.D., Michael R.

**TITLE**  
**JOURNAL**  
**COMMENT**

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 On Sep 12, 1996 this sequence version replaced gi:1393355.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: Stratagene, Inc., David B. Krizman,  
 Ph.D.  
 CDNA Library Arraying: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Insert Length: 1661 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 87.

**FEATURES**  
 source  
 Location/Qualifiers  
 1..107  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1016157"  
 /clone\_lib="NCI-CGAP\_GC2"  
 /tissue\_type="germ cell tumor"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; cloned unidirectionally. Primer: Oligo dT. Bulk germ cell tumor. 5' adaptor sequence: 5' GAATTCGGCAGCAG 3' 3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'  
 Average insert size: 1.2 kb."  
 22 a 34 c 26 g 25 t

Query Match 0.3%; Score 86.2; DB 35; Length 107;  
 Best Local Similarity 87.9%; Pred. No. 0.9;  
 Matches 94; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 27544 TGGCATGTGCTGTAAATCCAGCTACTCAGAGGCTCAGGCAAGAAATGCTTGAACCC 27603  
 Db 107 TGGTGTGCTGTAAATCCAGCTACTCAGAGGCTCAGGCAAGAAATGCTTGAACCT 48

QY 27604 AGGAGCGGAGGTTGCAGTGAGCGGAATCCGCCACTGCACTCCAG 27650  
 Db 47 GGGAGCGGAGGCTTGCAGTGAGCTGAGATTGAGCCACTGCACTCCAG 1

**RESULT** 11  
**AA828124/c**  
**LOCUS**  
**DEFINITION**  
**ACCSSION**  
**VERSION**  
**KEYWORDS**  
**SOURCE**  
**ORGANISM**  
**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**  
**COMMENT**

AA828124 107 bp mRNA EST 20-FEB-1998  
 od71a07.s1 NCI-CGAP\_Ov2 Homo sapiens cDNA clone IMAGE:1373364  
 similar to contains Alu repetitive element;contains element MBR22  
 repetitive element ;, mRNA sequence.  
 AA828124  
 AA828124.1 GI:2900487  
 EST.  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 107)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 On Jan 17, 1998 this sequence version replaced gi:1899815.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher A. Moskaluk, M.D., Michael R.

Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: David B. Krizman, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 93.

#### FEATURES

source  
1..107  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1373364"  
/clone\_lib="NCI\_CGAP\_Ov2"  
/sex="female"  
/tissue\_type="ovary"  
/lab\_host="DH10B"  
/note="Vector: pAMP10; mRNA made from invasive ovarian  
tumor, cDNA made by oligo-dT priming. Non-directionally  
cloned. Size-selected on agarose gel, average insert size  
600 bp. Reference: Krizman et al. (1996) Cancer Research  
56:5380-5383."

BASE COUNT 30 a 23 c 38 g 16 t

#### ORIGIN

Query Match 0.3%; Score 86.4; DB 39; Length 107;  
Best Local Similarity 89.4%; Pred. No. 0.86;  
Matches 93; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
QY 22323 TCCGCTCCCGGATTCACGCCATTCCTCCGCTCAACCTCCGAGTAGTGGGACACAG 22382  
Db 107 TCGGCTCCCGGGTTCACGCCATTCCTCCGCTCAACCTCCGAGTAGTGGGACACAG 48  
QY 22383 GCGCCGCCACACGCCGCCAGCTAATTTTGTATTTTAGTAGA 22426  
Db 47 GTGCCGCCATTACGCCCTGCTAATTTTGTATTTTAGTAGCA 4

#### RESULT 12

AQ003188  
LOCUS 110 bp DNA GSS 14-APR-1999  
DEFINITION RPC11-1D10.TPN RPC1-11 Homo sapiens genomic clone RPC1-11-ID10,  
genomic survey sequence.  
ACCESSION AQ003188  
VERSION AQ003188.1 GI:3030392  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 110)  
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,  
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and  
Venter,J.C.  
TITLE Use of BAC End Sequences for Sequence-Ready Map Building (1998)  
JOURNAL Unpublished (1998)  
COMMENT Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdamas@tigr.org  
Clones are derived from the human BAC library RPC1-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from  
Research Genetics ([inforesgen.com](http://inforesgen.com)). BAC end search page:  
[http://www.tigr.org/tdb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html)  
Seq primer: SP6

#### Class: BAC ends.

#### FEATURES

source  
1..110  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="GDB:7500081"  
/db\_xref="taxon:9606"  
/clone="RPC1-11-ID10"  
/clone\_lib="RPC1-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPC111 Human Male BAC Library"

BASE COUNT 22 a 27 c 26 g 35 t

#### ORIGIN

Query Match 0.3%; Score 86.6; DB 94; Length 110;  
Best Local Similarity 87.2%; Pred. No. 0.81;  
Matches 95; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
QY 22408 TTTTGTATTTTAGTAGACAGAGGGTTTCACCGTGTAGCCGGGATGCTCGATCTCC 22467  
Db 2 TTTTGTATTTTAGTAGACAGAGGGTTTACCATTTTGGCCAGGATGCCGATCTCT 61  
QY 22468 TGACCTCATGATCTGCCACCTCAGCCTCCCAAGTCTAGGATCACAG 22516  
Db 62 TGACCTCATGATCCACCTGCCGAGCCTCCCAAGTCTGGGATTACAG 110

#### RESULT 13

AA654562  
LOCUS 102 bp mRNA EST 04-NOV-1997  
DEFINITION nt75f10.s1 NCI\_CGAP\_Pr3 Homo sapiens cDNA clone IMAGE:1204363  
similar to contains Alu repetitive element; contains element MER22  
repetitive element ;, mRNA sequence.  
ACCESSION AA654562  
VERSION AA654562.1 GI:2590716  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 102)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1393451.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: David B. Krizman, Ph.D.  
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Seq primer: -40ml3 fwd. ET from Amersham.

#### FEATURES

source  
1..102  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1204363"  
/clone\_lib="NCI\_CGAP\_Pr3"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"  
/note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from









GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2000, 03:08:43 ; Search time 372.76 seconds  
(without alignments)  
10112.932 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_112000\_141000  
Perfect score: 29001  
Sequence: 1 TGTACTGTGCTACCTCTG.....CTATAAAGTGAACAACCC 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 374504

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/PCUS\_COMB.seq:\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78.4	0.3	105	4	US-08-481-658B-65
2	78.4	0.3	105	4	US-08-477-504A-65
3	78.4	0.3	105	4	US-08-486-756A-65
4	78.4	0.3	105	4	US-08-485-862B-65
5	78.4	0.3	105	5	US-08-787-739-65
6	68	0.2	105	4	US-08-481-658B-65
7	68	0.2	105	4	US-08-477-504A-65
8	68	0.2	105	4	US-08-486-756A-65
9	68	0.2	105	4	US-08-485-862B-65
10	68	0.2	105	5	US-08-787-739-65
11	60.4	0.2	78	3	US-08-454-557C-70
12	60.4	0.2	78	4	US-08-340-426D-70
13	60.4	0.2	78	4	US-08-450-673C-70
14	60.4	0.2	78	6	PCT-US95-17111A-70
15	58.8	0.2	78	3	US-08-454-557C-70
16	58.8	0.2	78	4	US-08-340-426D-70
17	58.8	0.2	78	4	US-08-450-673C-70
18	58.8	0.2	78	6	PCT-US95-17111A-70
19	59	0.2	84	3	US-08-454-557C-91
20	59	0.2	84	4	US-08-340-426D-91
21	59	0.2	84	4	US-08-450-673C-91
22	59	0.2	84	6	PCT-US95-17111A-91
23	53.2	0.2	76	3	US-08-454-557C-69
24	53.2	0.2	76	4	US-08-340-426D-69
25	53.2	0.2	76	4	US-08-450-673C-69
26	53.2	0.2	76	6	PCT-US95-17111A-69
27	53.4	0.2	85	3	US-08-454-557C-92

28	53.4	0.2	85	4	US-08-340-426D-92	Sequence 92, Appl
29	53.4	0.2	85	4	US-08-450-673C-92	Sequence 92, Appl
30	53.4	0.2	85	6	PCT-US95-17111A-92	Sequence 92, Appl
31	51.8	0.2	84	3	US-08-454-557C-91	Sequence 91, Appl
32	51.8	0.2	84	4	US-08-340-426D-91	Sequence 91, Appl
33	51.8	0.2	84	4	US-08-450-673C-91	Sequence 91, Appl
34	51.8	0.2	84	6	PCT-US95-17111A-91	Sequence 91, Appl
35	50.4	0.2	60	3	US-08-454-557C-57	Sequence 57, Appl
36	50.4	0.2	60	4	US-08-340-426D-57	Sequence 57, Appl
37	50.4	0.2	60	4	US-08-450-673C-57	Sequence 57, Appl
38	50.4	0.2	60	6	PCT-US95-17111A-57	Sequence 57, Appl
39	50	0.2	76	3	US-08-454-557C-69	Sequence 69, Appl
40	50	0.2	76	4	US-08-340-426D-69	Sequence 69, Appl
41	50	0.2	76	4	US-08-450-673C-69	Sequence 69, Appl
42	50	0.2	76	6	PCT-US95-17111A-69	Sequence 69, Appl
43	48.8	0.2	60	3	US-08-454-557C-60	Sequence 60, Appl
44	48.8	0.2	60	3	US-08-454-557C-60	Sequence 60, Appl
45	48.8	0.2	60	4	US-08-340-426D-60	Sequence 60, Appl

ALIGNMENTS

RESULT 1  
US-08-481-658B-65  
; Sequence 65, Application US/08481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,658B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.3E  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-481-658B-65

Query Match 0.3%; Score 78.4; DB 4; Length 105;  
Best Local Similarity 84.6%; Pred. No. 2.4e-08;

Matches 88; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 21613 TTTTGTATTTTGTAGTAAGAGCGGGTTTACCATGTTGGTCAGGCTGGTCTCCAACCTCC 21672

Db 2 TTTTGTATCTTTAGTAGAGAGCGGTTTACCATATTGGCCAGGCTGCTCTCAAACTCC 61

Qy 21673 TGACCTCATGATCTGCCACCTTGCCCTCCCAAAAGTGTGGGAT 21716

Db 62 TGACCTTGATCCACCAGCCTCGSCCTCCCAAAAGTGTGGGAT 105

## RESULT 2

US-08-477-504A-65

; Sequence 65, Application US/08477504A

; Patent No. 5972353

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

; CITY: Tiburon

; STATE: California

; COUNTRY: USA

; ZIP: 94920

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/477,504A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/260,190

; FILING DATE: 15-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lauder, Leona L.

; REGISTRATION NUMBER: 30,863

; REFERENCE/DOCKET NUMBER: D-0021.3D

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-435-2034

; TELEFAX: 415-435-0727

; INFORMATION FOR SEQ ID NO: 65:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 105 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-477-504A-65

Query Match 0.3%; Score 78.4; DB 4; Length 105;

Best Local Similarity 84.6%; Pred. No. 2.4e-08;

Matches 88; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 21613 TTTTGTATTTTGTAGTAAGAGCGGGTTTACCATGTTGGTCAGGCTGGTCTCCAACCTCC 21672

Db 2 TTTTGTATCTTTAGTAGAGAGCGGTTTACCATATTGGCCAGGCTGCTCTCAAACTCC 61

Qy 21673 TGACCTCATGATCTGCCACCTTGCCCTCCCAAAAGTGTGGGAT 21716

Db 62 TGACCTTGATCCACCAGCCTCGSCCTCCCAAAAGTGTGGGAT 105

## RESULT 3

US-08-486-756A-65

; Sequence 65, Application US/08486756A

; Patent No. 5981711

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

; CITY: Tiburon

; STATE: California

; COUNTRY: USA

; ZIP: 94920

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/486,756A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/260,190

; FILING DATE: 15-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lauder, Leona L.

; REGISTRATION NUMBER: 30,863

; REFERENCE/DOCKET NUMBER: D-0021.3C

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-435-2034

; TELEFAX: 415-435-0727

; INFORMATION FOR SEQ ID NO: 65:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 105 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-486-756A-65

Query Match 0.3%; Score 78.4; DB 4; Length 105;

Best Local Similarity 84.6%; Pred. No. 2.4e-08;

Matches 88; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 21613 TTTTGTATTTTGTAGTAAGAGCGGGTTTACCATGTTGGTCAGGCTGGTCTCCAACCTCC 21672

Db 2 TTTTGTATCTTTAGTAGAGAGCGGTTTACCATATTGGCCAGGCTGCTCTCAAACTCC 61

Qy 21673 TGACCTCATGATCTGCCACCTTGCCCTCCCAAAAGTGTGGGAT 21716

Db 62 TGACCTTGATCCACCAGCCTCGSCCTCCCAAAAGTGTGGGAT 105

## RESULT 4

US-08-485-862B-65

; Sequence 65, Application US/08485862B

; Patent No. 5989838

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

```
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,862B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/477,504
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-0727
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-485-862B-65

Query Match 0.3%; Score 78.4; DB 4; Length 105;
Best Local Similarity 84.6%; Pred. No. 2.4e-08;
Matches 88; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 21613 TTTTGTATTTTGTAGTAAGACGGGTTTCACCATGTTGGTCAGGCTGCTCCAACTCC 21672
Db 2 TTTTACATCTTTAGTAGACAGGTTTCACCATATTGGCCAGGCTGCTCTCAAACTCC 61

QY 21673 TGACCTCATGATCTGCCACCTTGGCCTCCCAAGTCTGGGAT 21716
Db 62 TGACCTTGTGATCCACAGCCTCGGCCTCCCAAGTCTGGGAT 105

RESULT 5
US-08-787-739-65
; Sequence 65, Application US/08/787739
; Patent No. 6027887
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 96
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 369 Pine Street, Suite 610
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/787,739
; FILING DATE: 24-JAN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,049
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/486,756
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/477,504
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/481,658
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,862
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,863
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,077
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-981-0334
; TELEFAX: 415-981-0332
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-787-739-65

Query Match 0.3%; Score 78.4; DB 5; Length 105;
Best Local Similarity 84.6%; Pred. No. 2.4e-08;
Matches 88; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 21613 TTTTGTATTTTGTAGTAAGACGGGTTTCACCATGTTGGTCAGGCTGCTCCAACTCC 21672
Db 2 TTTTACATCTTTAGTAGACAGGTTTCACCATATTGGCCAGGCTGCTCTCAAACTCC 61

QY 21673 TGACCTCATGATCTGCCACCTTGGCCTCCCAAGTCTGGGAT 21716
Db 62 TGACCTTGTGATCCACAGCCTCGGCCTCCCAAGTCTGGGAT 105

RESULT 6
US-08-481-658B-65/C
; Sequence 65, Application US/08/481658B
; Patent No. 5955075
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA: US/08/481,658B
; APPLICATION NUMBER: US/08/481,658B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-481-658B-65

Query Match      0.2%; Score 68; DB 4; Length 105;
Best Local Similarity 84.0%; Pred. No. 4.3e-06;
Matches 89; Conservative 0; Mismatches 15; Indels 2; Gaps 1;

QY 15260 ATACCAGCAGCTTTGGGAGCGCGATGTGGTGGATCACCCTGAGTCAGGAGTTTGAGACCA 15319
|| ||||||||||||||||||| ||||||||| ||||||||||||||||| ||
Db 105 ATCCAGCAGCTTTGGGAGCGCGAGCTGGTGATCAC--AAGGTCAGGAGTTTGAGAGCA 48

QY 15320 GACTGGCCCAACATGTTGAAACCTCATCTCTAGTAAATAACAAAA 15365
| ||||||| ||||||||| ||||||| ||||| ||| ||| |||
Db 47 GCCTGGCCCAATATGTTGAAACCCCTGCTCTACTAAGATGTAAAAA 2

RESULT 7
US-08-477-504A-65/c
; Sequence 65, Application US/08/477504A
; Patent No. 5972353
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,504A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-477-504A-65

Query Match      0.2%; Score 68; DB 4; Length 105;
Best Local Similarity 84.0%; Pred. No. 4.3e-06;
Matches 89; Conservative 0; Mismatches 15; Indels 2; Gaps 1;

QY 15260 ATACCAGCAGCTTTGGGAGCGCGATGTGGTGGATCACCCTGAGTCAGGAGTTTGAGACCA 15319
|| ||||||||||||||||||| ||||||||| ||||||||||||||||| ||
Db 105 ATCCAGCAGCTTTGGGAGCGCGAGCTGGTGATCAC--AAGGTCAGGAGTTTGAGAGCA 48

QY 15320 GACTGGCCCAACATGTTGAAACCTCATCTCTAGTAAATAACAAAA 15365
| ||||||| ||||||||| ||||||| ||||| ||| ||| |||
Db 47 GCCTGGCCCAATATGTTGAAACCCCTGCTCTACTAAGATGTAAAAA 2

RESULT 8
US-08-486-756A-65/c
; Sequence 65, Application US/08/486756A
; Patent No. 5981711
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,756A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-486-756A-65

Query Match 0.2%; Score 68; DB 4; Length 105;  
Best Local Similarity 84.0%; Pred. No. 4.3e-06;  
Matches 89; Conservative 0; Mismatches 15; Indels 2; Gaps 1;

QY 15260 ATACGAGCACTTTGGAGGCCGATGGTGGATCACCTGAGGTGAGGAGTTTGAGACCA 15319  
|| |||||  
DB 105 ATCCGAGCACTTTGGAGGCCGAGGCTGGTGGATCAC--AAGGTCAGGAGTTTGAGAGCA 48  
|| |||||  
QY 15320 GACTGGCCAACTGTGAACCTCATCTCTAGTAAAAATACAAAA 15365  
|| |||||  
DB 47 GCCTGCCAATATGGTGAACCCCTGCTCTACTAAAGATGTAATAA 2  
|| |||||

RESULT 9  
US-08-485-862B-65/c  
; Sequence 65, Application US/08485862B  
; Patent No. 5989838  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485.862B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/477,504  
; FILING DATE: 07-JUN-1995  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.3D  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-485-862B-65

Query Match 0.2%; Score 68; DB 4; Length 105;  
Best Local Similarity 84.0%; Pred. No. 4.3e-06;  
Matches 89; Conservative 0; Mismatches 15; Indels 2; Gaps 1;

QY 15260 ATACGAGCACTTTGGAGGCCGATGGTGGATCACCTGAGGTGAGGAGTTTGAGACCA 15319  
|| |||||  
DB 105 ATCCGAGCACTTTGGAGGCCGAGGCTGGTGGATCAC--AAGGTCAGGAGTTTGAGAGCA 48  
|| |||||  
QY 15320 GACTGGCCAACTGTGAACCTCATCTCTAGTAAAAATACAAAA 15365  
|| |||||  
DB 47 GCCTGCCAATATGGTGAACCCCTGCTCTACTAAAGATGTAATAA 2  
|| |||||

RESULT 10  
US-08-787-739-65/c  
; Sequence 65, Application US/08787739  
; Patent No. 6027887  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 96  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 369 Pine Street, Suite 610  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/787,739  
; FILING DATE: 24-JAN-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,049  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/486,756  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/477,504  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/481,658  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,862  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,863  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/487,077  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-981-2034  
; TELEFAX: 415-981-0332  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-787-739-65





Search completed: June 19, 2000, 11:34:48  
Job time: 427218 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2000, 11:28:29 ; Search time 17971.8 Seconds  
(without alignments)  
-1569.789 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_140000\_169000  
Perfect score: 29001  
Sequence: 1 GGTTCGACAAAGGTGTCAA.....TCCTTCAGGAGTTACTTCTA 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 370290

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_bal.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pl1.\*
- 8: gb\_pl2.\*
- 9: gb\_pr1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_vi.\*
- 17: em\_fun.\*
- 18: em\_hum1.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_om.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pl.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_vi.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_bal.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*
- 44: gb\_htg6.\*

- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*
- 50: gb\_pl3.\*
- 51: gb\_pr5.\*
- 52: gb\_htg8.\*
- 53: gb\_htg9.\*
- 54: gb\_htg10.\*
- 55: gb\_htg11.\*
- 56: gb\_htg12.\*
- 57: gb\_htg13.\*
- 58: gb\_htg14.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	85.8	0.3	108	11	HSU67803 Human small
2	85.2	0.3	108	10	HSLDLRN2 X05250 Human LDL-r
3	80.6	0.3	107	9	HUMALCE162 M87924 Human carc1
4	79.2	0.3	104	9	HUMALCE272 M87899 Human carc1
5	78.2	0.3	103	9	HUMALCE221 M87896 Human carc1
c 6	77.6	0.3	108	10	HSLDLRN2 X05250 Human LDL-r
7	77.8	0.3	108	11	HSU67804 Human small
8	76.2	0.3	108	11	HSU67808 U67808 Human small
9	75.4	0.3	103	13	HS8IC8R X57789 Human seque
c 10	75	0.3	97	9	HUMLDLRA1 M14178 Human low d
c 11	75	0.3	107	9	HUMALCE162 M87924 Human carc1
c 12	74.4	0.3	103	13	HS8IC8R X57789 Human seque
c 13	74.6	0.3	108	10	HSLDLRO1 X05249 Human LDL-r
14	74.6	0.3	108	10	HSLDLRO2 X05251 Human LDL-r
15	74.4	0.3	110	11	HSU67807 U67807 Human small
16	74	0.3	91	13	HUMUT8164A L30244 Human STS U
c 17	72.4	0.2	101	10	S79560 HRX (intron
c 18	72.4	0.2	103	9	HUMALCE221 M87896 Human carc1
c 19	72.4	0.2	108	10	HSLDLI12 X05248 Human LDL-r
c 20	72.6	0.3	108	13	G43535 WIAF-2393-S
21	72.4	0.2	110	9	HUMALCE43 M87900 Human carc1
c 22	72	0.2	97	9	HUMLDLRA2 M14180 Human low d
c 23	71.8	0.2	110	11	HSU67807 U67807 Human small
c 24	71.4	0.2	97	9	HUMLDLROJ M14179 Human fam1
25	71	0.2	97	9	HUMLDLRA2 M14180 Human low d
c 26	70.8	0.2	100	13	HUMUT931A L31299 Human STS U
c 27	71	0.2	105	13	G32655 G32655 A009L30 Hum
28	70.2	0.2	97	9	HUMLDLRA1 M14178 Human low d
29	69.8	0.2	107	11	HSU67806 U67806 Human small
30	69.4	0.2	108	9	HUMDID03M5 D16965 Human HepG2
c 31	68.4	0.2	95	13	HUMUT8002B L30176 Human STS U
c 32	68.6	0.2	97	9	HUMLDLROJ M14179 Human fam1
33	68.6	0.2	108	10	HSLDLRO1 X05249 Human LDL-r
c 34	68.6	0.2	108	10	HSLDLRO2 X05251 Human LDL-r
35	67.6	0.2	90	9	HUMLDLRM M15365 Human low d
36	67.8	0.2	108	13	G43535 G43535 WIAF-2393-S
37	67.2	0.2	100	9	HUMGALNSA D45223 Human GALNS
c 38	66.8	0.2	95	10	HSSTHPKTB X66361 H.sapiens m
39	66	0.2	100	10	HSLAS27 X91545 H.sapiens D
40	66	0.2	100	13	HUMUT931A L31299 Human STS U
c 41	65.6	0.2	90	9	HUMLDLRFL K03555 Human low d
42	65.8	0.2	101	10	S79560 HRX (intron
c 43	65.4	0.2	84	5	AR051521 AR051521 Sequence
c 44	65.6	0.2	102	13	G32906 G32906 A009M09 Hum
45	65.6	0.2	109	13	G02323 G02323 human STS S

ALIGNMENTS

```
RESULT 1
HSU67803      108 bp      RNA      PRI      01-AUG-1997
LOCUS         Human small cytoplasmic ALU transcript.
DEFINITION    U67803
ACCESSION     U67803.1 GI:2289917
VERSION       ALU.
KEYWORDS      human.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE     1 (bases 1 to 108)
AUTHORS       Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE         cDNAs derived from primary and small cytoplasmic ALU (sALU)
               transcripts
JOURNAL       J. Mol. Biol. 271 (2), 222-234 (1997)
MEDLINE       97415756
REFERENCE     2 (bases 1 to 108)
AUTHORS       Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE         Direct Submission
JOURNAL       Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The
               Children's Hospital of Philadelphia, 1004F Abramson Research
               Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA
               Location/Qualifiers
FEATURES      source
               1..108
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="TSCALu2"
               /note="sCALu"
               /rpt_family="Alu"
               /rpt_type="dispersed"
               /rpt_type="dispersed"
               repeat_region
               1..108
               /note="sCALu"
               /db_xref="taxon:9606"
               /clone="TSCALu2"
               /note="sCALu"
               /rpt_family="Alu"
               /rpt_type="dispersed"
               /rpt_type="dispersed"
BASE COUNT    23 a 39 c 30 g 16 t
ORIGIN
Query Match   0.3%; Score 85.8; DB 11; Length 108;
Best Local Similarity 92.8%; Pred. No. 0.0008;
Matches 90; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Qy 26228 GCCTGTATCCACGACTTTGGGAGGCTGGTGGTGAATCAGAGTTCAGGAGATCAA 26287
Db 1 GCCTGTATCCACGACTTTGGGAGGCTGGTGGTGAATCAGAGTTCAGGAGATCGA 60
Qy 26288 GACCATCTGCCACATGTTGAACCCCGTCTCTAC 26324
Db 61 GACCATCTGGCTAACAGGTGAACCCCGTCTCTAC 97

RESULT 2
HSLDLN2      108 bp      DNA      PRI      20-MAY-1992
LOCUS         Human LDL-receptor gene Intron 14 fragment (normal gene).
DEFINITION    X05250
ACCESSION     X05250
VERSION       X05250.1 GI:34337
KEYWORDS      Alu repetitive sequence; low density lipoprotein receptor.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
               Primates; Catarrhini; Homiidae; Homo.
REFERENCE     1 (bases 1 to 108)
AUTHORS       Horsthemke,B., Beisiegel,U., Dunning,A., Hovinga,J.R.,
               Williamson,R. and Humphries,S.
TITLE         Unequal crossing-over between two alu-repetitive DNA sequences in
               the low-density-lipoprotein-receptor gene. A possible mechanism for
               the defect in a patient with familial hypercholesterolaemia
JOURNAL       Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE       87161901
COMMENT       See X05252 for deletion junction
               Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
               Location/Qualifiers
FEATURES      source
               1..108
```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
1..108
/note="intron XIV fragment"
BASE COUNT    28 a 23 c 39 g 18 t
ORIGIN
Query Match   0.3%; Score 85.2; DB 10; Length 108;
Best Local Similarity 87.7%; Pred. No. 0.00098;
Matches 93; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Qy 26334 AAAAATAGCCAGCATGGTAGCACATGCTGTATCCAGCTACTCAAGAGGCTGAGGC 26393
Db 3 AAAAATAGCCAGCGTGGTGGCAGGTGCTGTATCCAGCTACTCGGAGGCTGAGGC 62
Qy 26394 AGGGGAATTGCTTGAACCCGGGAGGTGGACATTCAGTGCAGTCTGAG 26439
Db 63 AGGAGAATTGCTTGAACCCAGGAGGAGGAGGTTCAGTGCAGTGCAGCGAG 108

RESULT 3
HUMALCE162   107 bp ss-RNA      PRI      15-APR-1994
LOCUS         Human carcinoma cell-derived Alu RNA transcript, clone CE162.
DEFINITION    M87924
ACCESSION     M87924.1 GI:174871
KEYWORDS      Alu repeat.
SOURCE        Homo sapiens male embryo carcinoma cDNA to other RNA.
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE     1 (bases 1 to 107)
AUTHORS       Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE         Alu RNA transcripts in human embryonal carcinoma cells. Model of
               post-transcriptional selection of master sequences
JOURNAL       J. Mol. Biol. (1992) In press
FEATURES      Location/Qualifiers
               1..107
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /cell_line="Ntera2D1"
               /dev_stage="embryo"
               /sex="male"
               /tissue_type="carcinoma"
BASE COUNT    28 a 30 c 35 g 14 t
ORIGIN
Query Match   0.3%; Score 80.6; DB 9; Length 107;
Best Local Similarity 86.4%; Pred. No. 0.0047;
Matches 89; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
Qy 2906 GGCAGGAGAAATCACTTGAACCTGGGAGGACAGAGGTTCAGTGGGAGGAGTGGGCCACT 2965
Db 5 GGCAGAGAAATGGCGTGAACCCGGGAGGCGGAGTTCAGTGGAGTGGGCCACT 64
Qy 2966 GCATCCAGCTGAGCACACAGCGAGACTCTGTCTCAAAAAA 3008
Db 65 GCACTCCAGCGCTGGGCGACAGCGAGACTCCGTCTCAAAAAA 107

RESULT 4
HUMALCE272   104 bp ss-RNA      PRI      15-APR-1994
LOCUS         Human carcinoma cell-derived Alu RNA transcript, clone CE272.
DEFINITION    M87899
ACCESSION     M87899.1 GI:174875
KEYWORDS      Alu repeat.
SOURCE        Homo sapiens male embryo carcinoma cDNA to other RNA.
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Primates; Catarrhini; Homiidae; Homo.
```



```
BASE COUNT      26 a      38 c      26 g      18 t
ORIGIN

Query Match      0.3%; Score 77.8; DB 11; Length 108;
Best Local Similarity 87.6%; Pred. No. 0.021;
Matches 85; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 26228 GCCTGTATCCAGCAGCTTTGGGAGGCTGAGGTGGTGAATCAGCAGGTCTCAGGAGATCAA 26287
|||||
Db 1 GCCTGTATCCAGCAGCTTTGGGAGGCTGAGGTGGTGAATCAGCAGGTCTCAGGAGATCAA 60
|||||

QY 26288 GACCATCTGCCCAACATGGTGAACCCCGTCTCTAC 26324
|||||
Db 61 GACCATCTGCCCAACATGGTGAACCCCGTCTCTCC 97
|||||

RESULT 8
HSU67808      108 bp      RNA      PRI      01-AUG-1997
LOCUS      Human small cytoplasmic Alu transcript.
ACCESSION      U67808
VERSION      U67808.1 GI:2289922
KEYWORDS      Alu.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 108)
AUTHORS      Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE      cDNAs derived from primary and small cytoplasmic Alu (scAlu)
transcripts
JOURNAL      J. Mol. Biol. 271 (2), 222-234 (1997)
MEDLINE      97413756
REFERENCE      2 (bases 1 to 108)
AUTHORS      Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE      Direct Submission
JOURNAL      Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The
Children's Hospital of Philadelphia, 1004F Abramson Research
Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA
FEATURES
source
1..108
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TscAlu7"
repeat_region 1..108
/notes="scAlu"
/rpt_family="Alu"
/rpt_type="dispersed"
BASE COUNT      22 a      37 c      28 g      21 t
ORIGIN

Query Match      0.3%; Score 76.2; DB 11; Length 108;
Best Local Similarity 86.6%; Pred. No. 0.021;
Matches 84; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 26228 GCCTGTATCCAGCAGCTTTGGGAGGCTGAGGTGGTGAATCAGCAGGTCTCAGGAGATCAA 26287
|||||
Db 1 GCCTGTATCCAGCAGCTTTGGGAGGCTGAGGTGGTGAATCAGCAGGTCTCAGGAGATCAA 60
|||||

QY 26288 GACCATCTGCCCAACATGGTGAACCCCGTCTCTAC 26324
|||||
Db 61 GACCATCTGCCCAACATGGTGAACCCCGTCTCTCC 97
|||||

RESULT 9
HS81C8R      103 bp      DNA      STS      05-SEP-1991
LOCUS      Human sequence tagged site 81C8R DNA from 19q13.
DEFINITION      X57789
ACCESSION      X57789
VERSION      X57789.1 GI:23938

KEYWORDS      STS; myotonic dystrophy.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 103)
AUTHORS      Aldridge,F.L.
TITLE      Direct Submission
JOURNAL      Submitted (12-FEB-1991) F.L. Aldridge, ICI Pharmaceuticals,
Alderley Park, Macclesfield, Cheshire, SK10 4TG, UK
REFERENCE      2 (bases 1 to 103)
AUTHORS      Butler,R., Riley,J.H., Ogilvie,D.J., Anand,R., Buxton,J.,
Davies,J., Johnson,K. and Markham,A.F.
TITLE      Two sequence-tagged sites defining the ends of a 380 kb YAC clone
from 19q13
JOURNAL      Nucleic Acids Res. 19 (17), 4787 (1991)
MEDLINE      91367697
COMMENT      See also X57788 for STS 81C8L.
FEATURES
source
1..103
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19q13"
/germline
/clone_lib="YAC library: ICI"
/clone="81C8"
BASE COUNT      29 a      28 c      23 g      22 t      1 others
ORIGIN

Query Match      0.3%; Score 75.4; DB 13; Length 103;
Best Local Similarity 87.2%; Pred. No. 0.028;
Matches 82; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 13498 CTGTAGAGATGGGTTTCGCCATGTTGGCCAGGTGTTCTCAAACTCTGACCTGAAGTG 13557
|||||
Db 1 CAGTAGAGATGGGTTTCGCCATGTTGGCCAGGTGTTCTCAAACTCTGACCTGAAGTG 60
|||||

QY 13558 TTCCACCCAGCTCGGCTGCCAAAGTGTGGATT 13591
|||||
Db 61 ATCCACCCAGCTCGGCTGCCAAAGTGTGGATT 94
|||||

RESULT 10
HUMDLRAL/C      97 bp      DNA      PRI      07-JAN-1995
LOCUS      Human low density lipoprotein receptor gene, intron 4 (partial).
DEFINITION      M14178
ACCESSION      M14178.1 GI:187097
KEYWORDS      low density lipoprotein receptor-1.
SEGMENT      1 of 2
SOURCE      Human white blood cell DNA.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 97)
AUTHORS      Hobbs,H.H., Brown,M.S., Goldstein,J.L. and Russell,D.W.
TITLE      Deletion of exon encoding cysteine-rich repeat of low density
lipoprotein receptor alters its binding specificity in a subject
with familial hypercholesterolemia
JOURNAL      J. Biol. Chem. 261 (28), 13114-13120 (1986)
MEDLINE      87008518
COMMENT      Analysis of the LDL-receptor gene of a patient with familial
hypercholesterolemia (FH) revealed the deletion of exon 5 resulting
from a homologous recombination between repetitive Alu sequences of
intron 4 and intron 5.
FEATURES
source
1..97
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="19p13.3"
<1..>97
/gene="LDLR"
intron
```

```
misc_feature
/note="LDL-receptor intron D; G00-119-362"
42..72
/gene="LDLR"
/note="deletion target sequence"
BASE COUNT      18 a   34 c   26 g   19 t
ORIGIN      Chromosome 19.

Query Match      0.3%; Score 75; DB 9; Length 97;
Best Local Similarity 89.0%; Pred. No. 0.032;
Matches 81; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 15704 GCCAGGTGTGGTCTATGCTGTAAATCCAGCACTTTGGGAGGCCAAGGCGCGGAT 15763
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 97 GCGGGTGACGGGCTCATGCTGCTGGAATCCCAAGCACTTTGGGAGGCCAAGGCGAGGAT 38

QY 15764 CACTTGAGCCAGGAGTTCACAGACCAGCGTG 15794
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 37 TGCTTTGAGCCAGGAGTTCAAGACCAGCGTG 7

RESULT 11
HUMALCE162/c HUMALCE162 107 bp ss-RNA PRI 15-APR-1994
LOCUS Human carcinoma cell-derived Alu RNA transcript, clone CE162.
DEFINITION M87924
ACCESSION M87924
VERSION M87924.1 GI:174871
KEYWORDS Alu repeat.
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 107)
AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
post-transcriptional selection of master sequences
JOURNAL J. Mol. Biol. (1992) In press
FEATURES
Location/Qualifiers
1..107
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="NTERa2D1"
/dev_stage="embryo"
/sex="male"
/tissue_type="carcinoma"
BASE COUNT      28 a   30 c   35 g   14 t
ORIGIN

Query Match      0.3%; Score 75; DB 9; Length 107;
Best Local Similarity 84.8%; Pred. No. 0.032;
Matches 84; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 18751 TTTTGTGAGTGGAGTCTTACTCTGTCTCAAGCTGGAGTCAGTGCCAGCAATCTCAGC 18810
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 107 TTTTGTGAGCGGAGTCTGCTCTGTCTGCGCCAGGCTGGAGTCAGTGCGGCGCATCTCGGC 48

QY 18811 TCATGCAACCTCTGCCCTCTGGGTTCACGCAATCTGC 18849
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 47 TCATGCAAGCTCCGCGCTCCCGGGTTCACGCCATCTTC 9

RESULT 12
HS81C8R/c HS81C8R 103 bp DNA STS 05-SEP-1991
LOCUS Human sequence tagged site 81C8R DNA from 19q13.
DEFINITION x57789
ACCESSION x57789
VERSION x57789.1 GI:23938
KEYWORDS STS; myotonic dystrophy.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 103)
AUTHORS Aldridge,F.L.
TITLE Direct Submision
JOURNAL Submitted (12-FEB-1991) F.L. Aldridge, ICI Pharmaceuticals,
Alderley Park, Macclesfield, Cheshire, SK10 4TG, UK
REFERENCE 2 (bases 1 to 103)
AUTHORS Butler,R., Riley,J.H., Ogilvie,D.J., Anand,R., Buxton,J.,
Davies,J., Johnson,K. and Markham,A.F.
TITLE Two sequence-tagged sites defining the ends of a 380 kb YAC clone
from 19q13
JOURNAL Nucleic Acids Res. 19 (17), 4787 (1991)
MEDLINE 91367697
COMMENT See also X57788 for STS 81C8L.
FEATURES
Location/Qualifiers
1..103
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19q13"
/germline
/clone_lib="YAC library: ICI"
/clone="81C8"
BASE COUNT      29 a   28 c   23 g   22 t   1 others
ORIGIN

Query Match      0.3%; Score 74.4; DB 13; Length 103;
Best Local Similarity 83.2%; Pred. No. 0.04;
Matches 84; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2741 TGCCTGTATATTCAGACACTCTGGAGGCGGAGTAGGTAGGTAGTGGAGGTCATGATT 2800
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 102 TGCCTATAATTTTCNGCACTTTGGAGGTGGAGTGGATCACTTAAGGTCAGGAGTT 43

QY 2801 CGAGACCACCTGGACACATAGTGAACCCCATCTCTACT 2841
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 42 CTTGACCAAGCTGGCCCAACATGGTGAACCCCATCTCTACT 2

RESULT 13
HS1DLRDL/c HS1DLRDL 108 bp DNA PRI 20-MAY-1992
LOCUS Human LDL-receptor mutated gene with intron 12 deletion junction.
DEFINITION X05249
ACCESSION X05249
VERSION X05249.1 GI:34335
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
Williamson,R. and Humphries,S.
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT *sources: hypercholesterol aemia
See X05248 for corresponding normal gene sequence
In the defective LDL-receptor gene the deletion occurred between two
alu-repetitive sequences, that are in the same direction, the
deletion eliminates exons 13 and 14 and changes the reading frame
of the resulting spliced mRNA.
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES
Location/Qualifiers
1..108
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="blood leukocytes from a patient with familial"
misc_feature 1..108
/note="deletion junction region intron 12/ Intron 15"
BASE COUNT      20 a   40 c   20 g   28 t
ORIGIN
```

Query Match 0.3%; Score 74.6; DB 10; Length 108;  
Best Local Similarity 81.9%; Pred. No. 0.037;  
Matches 86; Conservative 0; Mismatches 19; Indels 0; Gaps 0;  
Qy 26334 AAAAAATTAGCCAGCATGGTAGACATGCCCTGTAATCCAGCTACTCAAGAGGCTGAGGC 26393  
|||||  
Db 106 AAAAAATTAGCCAGCGTGGTGGCAGGTGCCCTGTAATCCAGCTACTCGGAGGCTGAGGC 47  
|||||  
Qy 26394 AGGGAATGCTTGAACCCGGAGGTGGACATTCAGTGTGCTGA 26438  
|||||  
Db 46 AGGAAATGTTTGAACCCAGGAGGAGGTTGTTGGTGAGCGCA 2

RESULT 14  
HSLDLRD2 108 bp DNA PRI 20-MAY-1992  
LOCUS Human LDL-receptor mutated gene with intron 14 deletion junction.  
ACCESSION X05251  
VERSION X05251.1 GI:34336  
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 108)  
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,  
Williamson,R. and Humphries,S.  
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in  
the low-density-lipoprotein-receptor gene. A possible mechanism for  
the defect in a patient with familial hypercholesterolaemia  
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)  
MEDLINE 87161901  
COMMENT \*source: hypercholesterol aemia  
See X05250 for corresponding normal gene sequence  
In the defective LDL-receptor gene the deletion occurred between two  
alu-repetitive sequences, that are in the same direction, the  
deletion eliminates exons 13 and 14 and changes the reading frame  
of the resulting spliced mRNA.  
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES  
source 1..108  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="blood leukocytes from a patient with familial"  
intron 1..108  
/note="intron XIV fragment"  
BASE COUNT 28 a 20 c 40 g 20 t  
ORIGIN

Query Match 0.3%; Score 74.6; DB 10; Length 108;  
Best Local Similarity 81.9%; Pred. No. 0.037;  
Matches 86; Conservative 0; Mismatches 19; Indels 0; Gaps 0;  
Qy 26334 AAAAAATTAGCCAGCATGGTAGACATGCCCTGTAATCCAGCTACTCAAGAGGCTGAGGC 26393  
|||||  
Db 3 AAAAAATTAGCCAGCGTGGTGGCAGGTGCCCTGTAATCCAGCTACTCGGAGGCTGAGGC 62  
|||||  
Qy 26394 AGGGAATGCTTGAACCCGGAGGTGGACATTCAGTGTGCTGA 26438  
|||||  
Db 63 AGGAAATGTTTGAACCCAGGAGGAGGTTGTTGGTGAGCGCA 107

RESULT 15  
HSU67807 110 bp RNA PRI 01-AUG-1997  
LOCUS Human small cytoplasmic Alu transcript.  
ACCESSION U67807  
VERSION U67807.1 GI:2289921  
KEYWORDS Alu.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 110)  
AUTHORS Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.  
TITLE cDNAs derived from primary and small cytoplasmic Alu (sAlu)  
transcripts  
JOURNAL J. Mol. Biol. 271 (2), 222-234 (1997)  
MEDLINE 97415756  
REFERENCE 2 (bases 1 to 110)  
AUTHORS Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.  
TITLE Direct Submission  
JOURNAL Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The  
Children's Hospital of Philadelphia, 1004F Abramson Research  
Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA  
FEATURES  
source 1..110  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="TscAlu6"  
repeat\_region 1..110  
/note="sAlu"  
/rpt\_family="Alu"  
BASE COUNT 26 a 39 c 24 g 21 t  
ORIGIN  
Query Match 0.3%; Score 74.4; DB 11; Length 110;  
Best Local Similarity 84.0%; Pred. No. 0.039;  
Matches 84; Conservative 0; Mismatches 16; Indels 0; Gaps 0;  
Qy 15723 GCCTGTATCCAGCCTTTGGGAGGCCAAGGCGGATCACTTGAAGCCAGGATTC 15782  
|||||  
Db 1 GCCTGTATCCAGCCTTTGGAGGCCAAGTGGTGGATCACTTGAGCCAGGATTC 60  
|||||  
Qy 15783 AAGACGAGCTGCCCAACATGGCAAAACCCCTGGCTCTACC 15822  
|||||  
Db 61 AAGACGAGCTGTCAACATGGTGAACCCCATCTTTCCC 100  
|||||  
Search completed: June 20, 2000, 02:22:00  
Job time: 481404 sec















```
RESULT 10
T25854
AC T25854 standard; cDNA to mRNA; 91 BP.
AD T25854;
DE Human gene signature HUMGS08084.
DE Human gene signature HUMGS08084.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1944; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 91 BP; 18 A; 22 C; 28 G; 18 T;

Query Match 0.2%; Score 61; DB 1; Length 91;
Best Local Similarity 78.7%; Pred. No. 2.2;
Matches 70; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Qy 15897 ATCACTTGAGCGGGGAGGCGAGAGGTTGACGTGAGCTGAGATTTGCGCAGCTACACTACAG 15956
Dbb 2 ATCACTTGAGCGCTAGAGGCGAGGCGTCAAGTGAGCTGAGATGCGCACTCTGCGCTCCAG 61

Qy 15957 CTGGGTGACAGAGAGAGATTTCTGTCTCA 15985
Dbb 62 CTCTGGTGACAGCGCTGAGANNCTGTCTCA 90

RESULT 11
T20927/c
ID T20927 standard; cDNA to mRNA; 103 BP.
AD T20927;
DE Human gene signature HUMGS02180.
DE Human gene signature HUMGS02180.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;

Query Match 0.2%; Score 61; DB 1; Length 91;
Best Local Similarity 78.7%; Pred. No. 2.2;
Matches 70; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Qy 15897 ATCACTTGAGCGGGGAGGCGAGAGGTTGACGTGAGCTGAGATTTGCGCAGCTACACTACAG 15956
Dbb 2 ATCACTTGAGCGCTAGAGGCGAGGCGTCAAGTGAGCTGAGATGCGCACTCTGCGCTCCAG 61

Qy 15957 CTGGGTGACAGAGAGAGATTTCTGTCTCA 15985
Dbb 62 CTCTGGTGACAGCGCTGAGANNCTGTCTCA 90

RESULT 12
V00420
ID V00420 standard; cDNA; 101 BP.
AD V00420;
DE 12-MAY-1998 (first entry)
DE 3' fragment of clone M97-2.
DE Human; secreted protein; molecular weight marker; genetic fingerprinting;
DE antibody production; nutritional supplement; therapy; neural tissue;
DE glioblastoma line TG98G; clone M97-2; ds.
OS Homo sapiens.
PN W09740069-A2.
PD 30-OCT-1997.
PR 14-APR-1997; U06134.
PR 19-APR-1996; US-635311.
PA (GEM) GENETICS INST INC.
PI Jacobs K, Lavellie ER, McCoy JM, Merberg D, Racie LA,
PI Spaulding V;
DR WPI; 97-535776/49.
PT Isolated nucleic acid clones from ATCC 98028 encode novel secreted
PT proteins - having many potential uses, e.g. as immunomodulators,
PT cell proliferation or differentiation inhibitors or haematopoiesis
PT regulators
PS Claim 25; Page 70; 114pp; English.
CC This sequence represents the 3' end of clone M97-2, which is a
CC polynucleotide of the invention. This sequence was isolated from a human
CC neural tissue (glioblastoma line TG98G) cDNA library. The polynucleotide,
CC which encodes a secreted protein, can be used, e.g. as a tissue or
CC molecular weight marker, in genetic fingerprinting, to raise anti-protein
CC or anti-DNA antibodies and in interaction trap assays. The protein can be
CC used to assay biological activity, raise antibodies for use in
CC immunoassays, as a marker, to identify inhibitors of its interactions and
CC as a nutritional supplement. It may also have a very wide range of
CC therapeutic and biological activities (no examples are given to support
CC this), e.g. cytokine or modulator of cell proliferation and
CC differentiation, immunostimulant or immunosuppressant, haematopoiesis
CC regulator, bone, cartilage, tendon, ligament and/or nerve tissue growth
```

```
CC stimulator, follicle inhibitor/stimulator, chemotactic/chemokinetic,
CC haemostatic, thrombolytic or anti-inflammatory agent, antimicrobial,
CC biorhythm, metabolism or behaviour modifier, anti-depressant or analgesic
CC or psoriasis treatative.
SQ Sequence 101 BP; 88 A; 0 C; 2 G; 11 T;

Query Match 0.2%; Score 60.6; DB 1; Length 101;
Best Local Similarity 79.1%; Pred. No. 2.5;
Matches 72; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Qy 7892 TGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAGAAAGAAAAAG 7951
D 11 TTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 70

Qy 7952 AAAAGAAAGACGATCAGCCATCAATA 7982
D 71 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 101

RESULT 13
T26213
ID T26213 standard; cDNA to mRNA; 103 BP.
AC T26213;
DE Human gene signature HUMGS08452.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PR 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (WATS)/ MATSUBARA K.
PA (OKUBO)/ OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 2029; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 103 BP; 33 A; 21 C; 25 G; 23 T;

Query Match 0.2%; Score 60.6; DB 1; Length 103;
Best Local Similarity 75.8%; Pred. No. 2.5;
Matches 75; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 20049 GATTGCTTAAGCCAGGAAATTAAGCTGAGGAGCCATGAGGCGCATGCACTCCA 20108
D 1 GATCATTGTAGTCCAGGAGTTGGTTTACAGTAGTATGATGCGCACCACTGCACTCCA 60

Qy 20109 GCCTGGGTCAGAGTAGACCCCTGTCTTAAAGAGATAA 20147
D 61 GCCTGGGCCACAGAGTAGAACAATGCTCTTTTANGAANA 99
```

## RESULT 14

```
V11595
ID V11595 standard; cDNA; 92 BP.
AC V11595;
DE Homo sapiens adult testes clone AX65_22 3' end.
KW adult; testes; cDNA library; clone AX65_22; anti-inflammatory;
KW therapeutic composition; autoimmune disease; immune; stimulation;
KW suppression; ds.
OS Homo sapiens.
PN WO9814576-A2.
PD 09-APR-1998.
PR 03-OCT-1997; UI8007.
PR 04-OCT-1996; US-726237.
PA (GEMY ) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 98-240082/21.
PT Nucleic acids encoding novel secreted proteins - useful as, e.g.
PT anti-inflammatory, immuno-stimulatory or suppressing agents
PS Disclosure; Page 63; 110pp; English.
CC The sequence is that of an isolated polynucleotide which may
CC be of use in the production of therapeutic compositions for
CC treating or ameliorating a medical condition in a mammal. Such
CC compositions may be used for, e.g. research purposes as markers for
CC tissues, molecular weight markers for gels, primers or probes, for
CC nutrition as carbon, nitrogen or carbohydrate source. They can also be
CC used as a cytokine for cell proliferation and differentiation activity,
CC as immune stimulants or suppressors, e.g. for viral, bacterial or fungal
CC infections, for autoimmune diseases such as multiple sclerosis or
CC systemic lupus erythematosus, to regulate haematopoiesis, for tissue
CC growth, as an activator or inhibitor, or as a chemotactic or
CC chemokinetic, haemostatic and thrombocytic, receptor/ligand,
CC anti-inflammatory or tumour inhibitor agents.
SQ Sequence 92 BP; 92 A; 0 C; 0 G; 0 U;

Query Match 0.2%; Score 60.2; DB 1; Length 92;
Best Local Similarity 79.8%; Pred. No. 2.8;
Matches 71; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 7898 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAGAAAGAAAAAGAAAAAG 7957
D 1 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 60

Qy 7958 AAAGAAACGATCAGCCATCAATCAACA 7986
D 61 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 89

RESULT 15
V61480
ID V61480 standard; cDNA; 92 BP.
AC V61480;
DE Human secreted protein bk95_3 3' cDNA.
KW Secreted protein; human; bk95_3; ds.
OS Homo sapiens.
PN WO9841533-A2.
PD 24-SEP-1998.
PR 19-MAR-1998; U05474.
PR 18-MAR-1998; US-040963.
PR 19-MAR-1997; US-820493.
PA (GEMY ) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 98-521163/44.
PT New polynucleotide(s) encoding secreted human proteins - derived
PT from human foetal kidney, adult testes and adult or foetal brain
PT cDNA libraries
PS Disclosure; Page 71; 112pp; English.
CC This is the 3' polyA sequence of a full-length cDNA clone,
```







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2000, 06:28:49 ; Search time 8512.45 seconds  
(without alignments)  
13808.895 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_140000\_169000  
Perfect score: 29001  
Sequence: 1 GGTTTGACAAAGGTGTCAA.....TCCTTCAGAGTACTTCTA 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 156056

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

- EST:\*
- 1: em\_est1:\*
  - 2: em\_est2:\*
  - 3: em\_est3:\*
  - 4: em\_est4:\*
  - 5: em\_est5:\*
  - 6: em\_est6:\*
  - 7: em\_est7:\*
  - 8: em\_est8:\*
  - 9: em\_est9:\*
  - 10: em\_est10:\*
  - 11: em\_est11:\*
  - 12: em\_est12:\*
  - 13: em\_est13:\*
  - 14: em\_est14:\*
  - 15: em\_est15:\*
  - 16: em\_est16:\*
  - 17: em\_est17:\*
  - 18: em\_est18:\*
  - 19: em\_est19:\*
  - 20: gb\_est1:\*
  - 21: gb\_est2:\*
  - 22: gb\_est3:\*
  - 23: gb\_est4:\*
  - 24: gb\_est5:\*
  - 25: gb\_est6:\*
  - 26: gb\_est7:\*
  - 27: gb\_est8:\*
  - 28: gb\_est9:\*
  - 29: gb\_est10:\*
  - 30: gb\_est11:\*
  - 31: gb\_est12:\*
  - 32: gb\_est13:\*
  - 33: gb\_est14:\*
  - 34: gb\_est15:\*
  - 35: gb\_est16:\*
  - 36: gb\_est17:\*
  - 37: gb\_est18:\*
  - 38: gb\_est19:\*
  - 39: gb\_est20:\*
  - 40: gb\_est21:\*
  - 41: gb\_est22:\*
  - 42: gb\_est23:\*
  - 43: gb\_est24:\*
  - 44: gb\_est25:\*

- 45: gb\_est26:\*
- 46: gb\_est27:\*
- 47: gb\_est28:\*
- 48: gb\_est29:\*
- 49: gb\_est30:\*
- 50: gb\_est31:\*
- 51: gb\_est32:\*
- 52: em\_est20:\*
- 53: em\_est21:\*
- 54: em\_est22:\*
- 55: em\_est23:\*
- 56: em\_est24:\*
- 57: em\_est25:\*
- 58: em\_est26:\*
- 59: gb\_est33:\*
- 60: gb\_est34:\*
- 61: gb\_est35:\*
- 62: gb\_est36:\*
- 63: gb\_est37:\*
- 64: gb\_est38:\*
- 65: em\_est27:\*
- 66: em\_est28:\*
- 67: em\_est29:\*
- 68: em\_est30:\*
- 69: gb\_est39:\*
- 70: gb\_est40:\*
- 71: gb\_est41:\*
- 72: gb\_est42:\*
- 73: gb\_est43:\*
- 74: gb\_est44:\*
- 75: em\_est31:\*
- 76: em\_est32:\*
- 77: em\_est33:\*
- 78: em\_est34:\*
- 79: gb\_est45:\*
- 80: gb\_est46:\*
- 81: gb\_est47:\*
- 82: gb\_gss1:\*
- 83: gb\_gss2:\*
- 84: gb\_gss3:\*
- 85: gb\_gss4:\*
- 86: em\_gss1:\*
- 87: em\_gss2:\*
- 88: em\_gss3:\*
- 89: em\_gss4:\*
- 90: gb\_gss5:\*
- 91: gb\_gss6:\*
- 92: gb\_gss7:\*
- 93: gb\_gss8:\*
- 94: gb\_gss9:\*
- 95: em\_gss5:\*
- 96: em\_gss6:\*
- 97: em\_gss7:\*
- 98: em\_gss8:\*
- 99: em\_gss9:\*
- 100: em\_gss10:\*
- 101: em\_gss11:\*
- 102: gb\_gss10:\*
- 103: gb\_gss11:\*
- 104: em\_gss12:\*
- 105: gb\_gss12:\*
- 106: gb\_gss13:\*
- 107: gb\_gss14:\*
- 108: gb\_gss15:\*
- 109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result % Query

No.	Score	Match	Length	DB	ID	Description
c 1	93.8	0.3	109	30	AA243009	AA243009 zr25h02.s
c 2	92.2	0.3	106	37	AA703692	AA703692 ag81a10.r
c 3	91.6	0.3	106	105	AQ264176	AQ264176 CITBI-El-
c 4	90.8	0.3	110	106	AQ386882	AQ386882 RPC111-13
c 5	87.4	0.3	105	105	AQ282107	AQ282107 RPC111-94
c 6	86	0.3	103	38	AA807640	AA807640 nx08b05.s
c 7	85.4	0.3	103	38	AQ535244	AQ535244 RPC1-11-3
c 8	85	0.3	110	39	AA897366	AA897366 am06h02.s
c 9	85	0.3	110	94	AQ003188	AQ003188 RPC111-1D
c 10	84.6	0.3	107	35	AA565533	AA565533 nk42b11.s
c 11	84.4	0.3	109	30	AA244173	AA244173 nc05h06.s
c 12	82.6	0.3	105	28	AA078003	AA078003 7H12D08 C
c 13	82.8	0.3	110	30	AA244245	AA244245 nc07a04.s
c 14	82.8	0.3	110	106	AQ386882	AQ386882 RPC111-13
c 15	82.2	0.3	103	35	AA570438	AA570438 nk63g02.s
c 16	82	0.3	103	84	B48914	B48914 RPC111-4A12
c 17	82.2	0.3	103	108	AQ334922	AQ334922 RPC1-11-3
c 18	82	0.3	106	63	AI991750	AI991750 wt48601.x
c 19	81.8	0.3	109	94	AQ029690	AQ029690 RPC111-41
c 20	82	0.3	110	32	AA369482	AA369482 EST80906
c 21	81.2	0.3	104	108	AQ544583	AQ544583 CITBI-El-
c 22	81.2	0.3	105	109	AQ637292	AQ637292 RPC1-11-4
c 23	81.2	0.3	106	30	AA250812	AA250812 zs06a05.s
c 24	81.4	0.3	107	35	AA583252	AA583252 nn41e04.s
c 25	81.4	0.3	108	84	B65160	B65160 CIT-HSP-201
c 26	80.6	0.3	103	108	AQ535244	AQ535244 RPC1-11-3
c 27	80.8	0.3	106	38	AA812141	AA812141 ob48h02.s
c 28	80.6	0.3	110	29	AA177157	AA177157 nc02g07.s
c 29	80.2	0.3	101	35	AA583697	AA583697 nn58f10.s
c 30	80.2	0.3	106	30	AA250812	AA250812 zs06a05.s
c 31	80.2	0.3	109	84	B17434	B17434 345K2.TVB C
c 32	80.4	0.3	110	30	AA244245	AA244245 nc07a04.s
c 33	79.8	0.3	101	33	AA381369	AA381369 EST94442
c 34	79.8	0.3	102	84	B48088	B48088 RPC111-4N6.
c 35	79.8	0.3	103	108	AQ584425	AQ584425 RPC1-11-4
c 36	80	0.3	104	105	AQ321855	AQ321855 RPC111-11
c 37	80	0.3	107	33	AA385808	AA385808 EST99495
c 38	80.2	0.3	108	32	AA370029	AA370029 EST81584
c 39	80.2	0.3	109	84	B17434	B17434 345K2.TVB C
c 40	79.6	0.3	106	20	T55212	T55212 yb43g11.sl
c 41	79.8	0.3	107	62	AI933497	AI933497 wm74d02.x
c 42	79.8	0.3	109	94	AQ029690	AQ029690 RPC111-41
c 43	79.4	0.3	105	61	AI832832	AI832832 at72g09.x
c 44	79.4	0.3	106	44	AI249096	AI249096 qb73g09.x
c 45	79.4	0.3	106	108	AQ544957	AQ544957 CITBI-El-

## ALIGNMENTS

```

RESULT 1
AA243009/c  AA243009  109 bp  mRNA  EST  11-MAR-1998
LOCUS      zr25h02.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens
DEFINITION cDNA clone IMAGE:664467 3' similar to contains Alu repetitive
            element;contains element LTR1 repetitive element ;, mRNA sequence.

ACCESSION  AA243009.1  GI:1873869
VERSION     EST.
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE   1 (bases 1 to 109)
            Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Willson,R.
            WashU-NCI human EST Project
            Unpublished (1997)
TITLE       WashU-NCI human EST Project
JOURNAL
COMMENT     On Dec 3, 1996 this sequence version replaced gi:1126869.

```

```

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1127 Std Error: 0.00
Seq primer: -41m13 fwd. ET from Amersham
High quality sequence stop: 102.
Location/Qualifiers
1. .109
/organism="Homo sapiens"
/db_xref="GDB:5426481"
/db_xref="taxon:9606"
/clone="IMAGE:664467"
/clone_lib="Stratagene NT2 neuronal precursor 937230"
/tissue_type="neuroepithelial cells"
/dev_stage="Ntera-2 neuroepithelial cells"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: brain; Vector: pBluescript SK-; Site:1:
ECORI; Site:2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Uninduced, exponentially growing neuroepithelial
cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT  19 a  30 c  30 g  30 t
ORIGIN

Query Match      0.3%; Score 93.8; DB 30; Length 109;
Best Local Similarity 93.3%; Pred. No. 0.11;
Matches 98; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 26225  CACGCCGTATCCAGCACATTTGGGAGGCTGAGTGGTGAATCAGAGTCAGGAGAT 26284
          |||||
Db 109  CACGCGTATATCCAGCACATTTGGGAGGCGCGGATGCGGATCAGGAGTAGGAGAT 50
          |||||

Qy 26285  CAAGCACCATCTGCCACACATGTTGACACCCGCTCTACTATAAA 26329
          |||||
Db 49  CAAGCACCATCTGGCTAACACGGTGAACACCCGCTCTACTATAAA 5
          |||||

RESULT 2
AA703692/c  AA703692  106 bp  mRNA  EST  24-DEC-1997
LOCUS      ag81a10.r1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone
DEFINITION IMAGE:1140858 5' similar to contains Alu repetitive element; , mRNA
            sequence.

ACCESSION  AA703692
VERSION     AA703692.1  GI:2713610
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE   1 (bases 1 to 106)
            Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Willson,R.
            WashU-NCI human EST Project
            Unpublished (1997)
TITLE       WashU-NCI human EST Project
JOURNAL
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1397630.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.

```

Seq primer: -28ml3 rev1 ET from Amer sham  
High quality sequence stop: 53.

# FEATURES

Location/Qualifiers  
1..106  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1140858"  
/clone\_lib="Stratagene hNT neuron (#937233)"  
/dev\_stage="hNT neurons"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Differentiated, post mitotic hNT neurons. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGCAGAG 3' -3' adaptor sequence: 5' CTCACGTTTCTTTTCTTTT 3"

BASE COUNT 19 a 29 c 29 g 29 t  
ORIGIN

Query Match 0.3%; Score 92.2; DB 37; Length 106;  
Best Local Similarity 92.4%; Pred. No. 0.17;  
Matches 97; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 26225 CACGCTCTAATCCAGCAGCTTTGGAGGCTGAGGTGGTGAATCAGCAGGTCAGGAGAT 26284  
|||||  
DB 106 CACGCTCTAATCCAGCAGCTTTGGAGGCTGAGGTGGTGAATCAGCAGGTCAGGAGAT 47  
QY 26285 CAGACCATCTGGCCACATGGTGAACCCGCTCTCTACTAAAA 26329  
|||||  
DB 46 CGAGACCATCTCTGTAACACGCGTGAACCTCGTCTACTAAAA 2

## RESULT 3

AQ264176 106 bp DNA GSS 27-OCT-1998  
LOCUS CITBI-EI-2509A2.TF CITBI-EI Homo sapiens genomic clone 2509A2,  
DEFINITION genomic survey sequence.  
ACCESSION AQ264176  
VERSION AQ264176.1 GI:3792743  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 106)  
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.  
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready  
JOURNAL Map Building  
COMMENT Other\_GSSs: CITBI-EI-2509A2.TR  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: M13-21  
Class: BAC ends.

# FEATURES

Location/Qualifiers  
1..106  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2509A2"  
/clone\_lib="CITBI-EI"  
/sex="male"  
/cell\_type="sperm"

/note="Vector: pBelobAC11; Site\_1: EcoRI; Site\_2: EcoRI;  
Caltech Human BAC Library D"  
BASE COUNT 25 a 30 c 34 g 17 t  
ORIGIN

Query Match 0.3%; Score 91.6; DB 105; Length 106;  
Best Local Similarity 91.5%; Pred. No. 0.2;  
Matches 97; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 26210 CCGGTGAGTCAGTCACGCTGTATATCCAGCAGCTTTGGAGGTCAGGTGGTGAATC 26269  
|||||  
DB 1 CCGGGCGCAGAGTCACGCTGTATATCCAGCAGCTTTGGAGAGCGAGCGGTGATC 60  
QY 26270 ACGAGTCAGGAGATCAAGACCATCTGCGCCCAACATGGTGAACCC 26315  
|||||  
DB 61 ACGAGTCAGGAGATCAAGACCGCTCTGCTAACAATGGTGAACCC 106

## RESULT 4

AQ386882/c 110 bp DNA GSS 21-MAY-1999  
LOCUS RPC111-134I4.TV RPCI-11 Homo sapiens genomic clone RPCI-11-134I4,  
DEFINITION genomic survey sequence.  
ACCESSION AQ386882  
VERSION AQ386882.1 GI:4357905  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 110)  
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.  
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
JOURNAL Map Building  
COMMENT Other\_GSSs: RPC111-134I4.TJ  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbs@tigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from  
Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html  
Seq primer: 17  
Class: BAC ends.

# FEATURES

Location/Qualifiers  
1..110  
/organism="Homo sapiens"  
/db\_xref="GDB:7551267"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-134I4"  
/clone\_lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPC111 Human Male BAC Library"  
BASE COUNT 26 a 26 c 38 g 20 t  
ORIGIN

Query Match 0.3%; Score 90.8; DB 106; Length 110;  
Best Local Similarity 89.1%; Pred. No. 0.25;  
Matches 98; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
QY 11717 GGGTTTCAACATGTTGGCCAGGCTGGTGTGAACCTCCTCAAGTCACCTGCC 11776

```

|||||
Db 110 GGGTTTCCACCTAGTGTCCAGCGTGGTCTTGAACCTCTGAGCTCAAGCGATCCACCTGCC 11
Qy 11777 TCAGCCTCACATAGTCTCGGAGTATACAGCGTGAGCCACCATGCTGGCC 11826
Db 50 TCAGCCTCCCAAGTACTTGGATTACAGCGTGAGCCCACTGCTCCCGGCC 1

RESULT 5
AQ282107 105 bp DNA GSS 27-APR-1999
LOCUS AQ282107.1 RP111-94B21.TJ RP11-11 Homo sapiens genomic clone RP11-11-94B21,
DEFINITION genomic survey sequence.
ACCESSION AQ282107
VERSION AQ282107.1 GI:3907976
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 105)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RP11-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.
FEATURES             Location/Qualifiers
     source
       1..105
         /organism="Homo sapiens"
         /db_xref="GDB:7535756"
         /db_xref="taxon:9606"
         /clone="RP11-11-94B21"
         /clone_lib="RP11-11"
         /sex="Male"
         /cell_type="Lymphocytes"
         /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RP111 Human Male BAC Library"
BASE COUNT          26 a 31 c 30 g 18 t
ORIGIN
Query Match          0.3%; Score 87.4; DB 105; Length 105;
Best Local Similarity 89.5%; Pred. No. 0.6;
Matches 94; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 26222 GCTACGCGCTGTATCCAGCAGCTTTGGAGCGTGGTGGTGAATCAGGAGTCAGGA 26281
|||||
Db 1 GCTACGCGCTGTATCCAGCAGCTCTGGGAGGCCAAGGTGGGTGGATCAGGAGGCATGA 60

Qy 26282 GATCAAGACCATCCCGCCACATCGTGAACCCCGTCTCTACTA 26326
|||||
Db 61 GTACGAGACCAAGCGCTGACCAACATGTTGAACCCCGTCTCTACTA 105

RESULT 6
AA807640/c 103 bp mRNA EST 05-MAR-1998
LOCUS AA807640.1 NCI-CGAP_GC3 Homo sapiens cDNA clone IMAGE:1255473 3'
DEFINITION similar to contains Alu repetitive element; mRNA sequence.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

ACCESSION AA807640.1 GI:2877108
VERSION AA807640.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 103)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2151346.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 774 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 87.
FEATURES             Location/Qualifiers
     source
       1..103
         /organism="Homo sapiens"
         /db_xref="taxon:9606"
         /clone="IMAGE:1255473"
         /clone_lib="NCI-CGAP_GC3"
         /tissue_type="pooled germ cell tumors"
         /lab_host="DH10B"
         /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled
germ cell tumors, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is not normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT          19 a 27 c 30 g 27 t
ORIGIN
Query Match          0.3%; Score 86; DB 38; Length 103;
Best Local Similarity 90.2%; Pred. No. 0.87;
Matches 92; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 26224 TCAGCGCTGTATCCAGCAGCTTTGGGAGGTGAGGTGGTGAATCAGGAGTCAGGAGA 26283
|||||
Db 103 TCACACCTGTATCCAGCAGCTTTGGGAGGCGGAGTGGAGGATCACAAGGTCTAGGAGA 44

Qy 26284 TCAGACCATCTCTGCCCAACATGGTGAACCCCGTCTCTACT 26325
|||||
Db 43 TCGAGACCATCTCTGGCTATCAGCGTGAACCCCATCTCTACT 2

RESULT 7
AQ535244 103 bp DNA GSS 18-MAY-1999
LOCUS AQ535244.1 RP11-11-317H22.TV RP11-11 Homo sapiens genomic clone
DEFINITION RP11-11-317H22, genomic survey sequence.
ACCESSION AQ535244
VERSION AQ535244.1 GI:4846934
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

```



```
source          1. .110
                /organism="Homo sapiens"
                /db_xref="GDB:7500081"
                /db_xref="taxon:9606"
                /clone="RPC1-11-1D10"
                /clone_lib="RPC1-11"
                /sex="Male"
                /cell_type="Lymphocytes"
                /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
                RPC111 Human Male BAC Library"
BASE COUNT      22 a 27 c 26 g 35 t
ORIGIN

Query Match      0.3%  Score 85;  DB 94;  Length 110;
Best Local Similarity 86.28; Pred. No. 1.1;
Matches 94; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 26230 CTGTAATCCAGCACTTTGGGAGGCTGAGGTGGGTGAATCACGAGGTCAAGATCAAGA 26289
Db 110 CTGTAATCCAGCACTTTGGGAGGCTGCGGCAGGTGGATCATGAGGTCAAGAGATCGGGA 51
Qy 26290 CCATCTCGGCCAACATGGTGAACCCCGTCTCTACTAAATACAAAAA 26338
Db 50 CCATCTCGGCCAACATGGTGAACCCCGTCTCTACTAAATACAAAAA 2

RESULT 10
AA565533/c
LOCUS
DEFINITION      nk42b11.s1 NCI-CGAP-GC2 Homo sapiens cDNA clone IMAGE:1016157 3'
                similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION      AA565533
VERSION        AA565533.1 GI:2337172
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 107)
AUTHORS       NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE         National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
                Tumor Gene Index
JOURNAL        Unpublished (1997)
COMMENT        On Sep 12, 1996 this sequence version replaced gi:1393355.
                Contact: Robert Strausberg, Ph.D.
                Tel.: (301) 496-1550
                Email: Robert.Strausberg@nih.gov
                Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
                Emmert-Buck, M.D., Ph.D.
                CDNA Library Preparation: Stratagene, Inc., David B. Krizman,
                Ph.D.
                CDNA Library Arraying: Greg Lennon, Ph.D.
                DNA Sequencing by: Washington University Genome Sequencing Center
                Clone distribution: NCI-CGAP clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1661 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 87.
Location/Qualifiers
1. .107
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1016157"
/clone_lib="NCI-CGAP-GC2"
/tissue_type="germ cell tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: Bluescript SK-; Site.1: EcoRI; Site.2:
XhoI; Cloned unidirectionally. Primer: Oligo dT. Bulk
germ cell tumor. 5' adaptor sequence: 5' GAATTCGGCAGCAG 3'
3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTTT 3'
```

```
BASE COUNT      22 a 34 c 26 g 25 t
ORIGIN

Query Match      0.3%  Score 84.6;  DB 35;  Length 107;
Best Local Similarity 86.9%; Pred. No. 1.2;
Matches 93; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 15950 TGGCACATGCTGTAAATCCAGCTACTTGGGAGGCTCAAGCACAAGATCAGTGAACCG 15909
Db 107 TGGTGTGCTGCTGTAAATCCAGCTACTCAGGAGGCTGAGGCAGAGAAATCAGTGAACCT 48
Qy 15910 GGGAGGCAGAGGTTGCAGTGCAGTGCAGATTTCGCCCATACACTACAG 15956
Db 47 GGGAGGCAGAGGTTGCAGTGCAGTGCAGATTTCGCCCATACACTACAG 1

RESULT 11
AA244173/c
LOCUS
DEFINITION      nc05h06.s1 NCI-CGAP_Pr1 Homo sapiens cDNA clone IMAGE:1007291
                similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION      AA244173
VERSION        AA244173.1 GI:1874876
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 109)
AUTHORS       NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE         National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
                Tumor Gene Index
JOURNAL        Unpublished (1997)
COMMENT        On Nov 29, 1993 this sequence version replaced gi:430513.
                Contact: Robert Strausberg, Ph.D.
                Tel.: (301) 496-1550
                Email: Robert.Strausberg@nih.gov
                Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
                M.D., Michael Emmert-Buck, M.D., Ph.D.
                CDNA Library Preparation: David B. Krizman, Ph.D.
                CDNA Library Arraying by: Genome Systems Inc., Greg Lennon, Ph.D.
                DNA Sequencing by: Washington University Genome Sequencing Center
                Clone distribution: NCI-CGAP clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -41m13 fwd. ET from Amersham
High quality sequence stop: 90.
Location/Qualifiers
1. .109
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1007291"
/clone_lib="NCI-CGAP_Pr1"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: PAMPl0; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected, histologically normal
prostate epithelial cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into pAMP10 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Krizman."
BASE COUNT      28 a 28 c 31 g 22 t
ORIGIN
```



AQ386882	LOCUS
AQ386882	110 bp DNA GSS 21-MAY-1999
RPCI11-13414.TV RPCI-11 Homo sapiens genomic clone RPCI-11-13414,	
genomic survey sequence.	
AQ386882	
AQ386882.1 GI:4357905	
GSS.	
Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
Eutheria; Primates; Catarrhini; Hominidae; Homo.	
1 (bases 1 to 110)	
Zhao,S.; Adams,M.D., Nierman,W., Malek,J., de Jong,P. and	
Venter,J.C.	
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready	
Map Building	
Unpublished (1997)	
Other_GSSs: RPCI11-13414.TJ	
Contact: Shaying Zhao, William Nierman, Mark Adams	
Department of Eukaryotic Genomics	
The Institute for Genomic Research	
9712 Medical Center Dr., Rockville, MD 20850	
Tel: 301 838 0200	
Fax: 301 838 0208	
Email: hbe@tigr.org	
Clones are derived from the human BAC library RPCI-11. For BAC	
library availability, please contact Pieter de Jong	
(pieterejong.med.buffalo.edu). Clones may be purchased from	
BACPAC Resources ( <a href="http://bacpac.med.buffalo.edu/ordering">http://bacpac.med.buffalo.edu/ordering</a> ) or from	
Research Genetics ( <a href="mailto:info@resgen.com">info@resgen.com</a> ). BAC end search page:	
<a href="http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html">http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html</a>	
Seq primer: 17	
Class: BAC ends.	
FEATURES	
Source	
Location/Qualifiers	
1..110	
/organism="Homo sapiens"	
/db_xref="GDB:7551267"	
/db_xref="taxon:9606"	
/clone_lib="RPCI-11-13414"	
/clone_id="RPCI-11"	
/sex="Male"	
/cell_type="Lymphocytes"	
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"	
BASE COUNT	26 a 26 c 38 g 20 t
ORIGIN	
Query Match	0.3%; Score 82.8; DB 106; Length 110;
Best Local Similarity	84.5%; Pred. No. 1.9;
Matches	93; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
YQ 15703 GGCCAGGTGGTGCATCGCTGAATCCAGCACATTTCGGAGGCCAAGCGCGGA 15762	
Db 1 GGCCGGAGCATGGCTCAGCGCTTAATCAAGTACTTTGGGAGGCTTGAGCGAGGTGA 60	







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2000, 11:34:48 ; Search time 372.06 Seconds  
(without alignments)  
10131.958 Million cell updates/sec

Title: US-08-852-495c-2\_copy\_140000\_169000  
Perfect score: 29001  
Sequence: 1 GCTTTTGACAAAGGTGTCAA.....TCCTTCAGGAGTTACTTCTA 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 374504

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/PTCUS\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	76.8	0.3	105	4	US-08-481-658B-65
C 2	76.8	0.3	105	4	US-08-477-504A-65
C 3	76.8	0.3	105	4	US-08-486-756A-65
C 4	76.8	0.3	105	4	US-08-485-862B-65
C 5	76.8	0.3	105	5	US-08-787-739-65
C 6	71.2	0.2	105	4	US-08-481-658B-65
C 7	71.2	0.2	105	4	US-08-477-504A-65
C 8	71.2	0.2	105	4	US-08-486-756A-65
C 9	71.2	0.2	105	4	US-08-485-862B-65
C 10	71.2	0.2	105	5	US-08-787-739-65
C 11	65.4	0.2	84	3	US-08-454-557C-91
C 12	65.4	0.2	84	4	US-08-340-426D-91
C 13	65.4	0.2	84	4	US-08-450-673C-91
C 14	65.4	0.2	84	6	PCT-US95-17111A-91
C 15	62.8	0.2	98	1	US-08-088-658-42
C 16	62.8	0.2	98	4	US-08-471-907A-42
C 17	60.4	0.2	78	3	US-08-454-557C-70
C 18	60.4	0.2	78	4	US-08-340-426D-70
C 19	60.4	0.2	78	4	US-08-450-673C-70
C 20	60.4	0.2	78	6	PCT-US95-17111A-70
C 21	59.6	0.2	84	2	US-08-664-596B-3
C 22	59.6	0.2	84	2	US-08-738-367-3
C 23	58.2	0.2	80	2	US-07-920-281C-25
C 24	57	0.2	75	3	US-08-454-557C-69
C 25	57	0.2	76	4	US-08-340-426D-69
C 26	57	0.2	76	4	US-08-450-673C-69
C 27	57	0.2	76	6	PCT-US95-17111A-69

28	55.8	0.2	69	2	US-08-702-344-7	Sequence 7, Appl
29	55.8	0.2	69	2	US-08-702-344-22	Sequence 92, Appl
C 30	56	0.2	85	3	US-08-454-557C-92	Sequence 22, Appl
C 31	56	0.2	85	4	US-08-340-426D-92	Sequence 92, Appl
C 32	56	0.2	85	4	US-08-450-673C-92	Sequence 92, Appl
C 33	56	0.2	85	6	PCT-US95-17111A-92	Sequence 92, Appl
C 34	55.6	0.2	78	3	US-08-454-557C-70	Sequence 70, Appl
C 35	55.6	0.2	78	4	US-08-340-426D-70	Sequence 70, Appl
C 36	55.6	0.2	78	4	US-08-450-673C-70	Sequence 70, Appl
C 37	55.6	0.2	78	6	PCT-US95-17111A-70	Sequence 70, Appl
C 38	55.2	0.2	60	3	US-08-454-557C-57	Sequence 57, Appl
C 39	55.2	0.2	60	4	US-08-340-426D-57	Sequence 57, Appl
C 40	55.2	0.2	60	4	US-08-450-673C-57	Sequence 57, Appl
C 41	55.2	0.2	60	6	PCT-US95-17111A-57	Sequence 57, Appl
C 42	55	0.2	84	3	US-08-454-557C-91	Sequence 91, Appl
C 43	55	0.2	84	4	US-08-340-426D-91	Sequence 91, Appl
C 44	55	0.2	84	4	US-08-450-673C-91	Sequence 91, Appl
C 45	55	0.2	84	6	PCT-US95-17111A-91	Sequence 91, Appl

ALIGNMENTS

RESULT 1  
US-08-481-658B-65/c  
; Sequence 65, Application US/08481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481.658B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.3E  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-481-658B-65

Query Match 0.3%; Score 76.8; DB 4; Length 105;  
Best Local Similarity 83.7%; Pred. No. 1.1e-06;

```
Matches 87; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 26235 ATCCAGCAGCTTTGGAGGCTGAGTGGTGAATCAGAGGTCAAGAGATCAAGACCATC 26294
|||||
Db 105 ATCCAGCAGCTTTGGAGGCGGAGGCTGGTGATCAAGGTCAAGGATTTGAGAGCAGC 46
|||||
QY 26295 CTGCCCAACATGGTGAACCCCGCTCTCTACTAAATAACAAAAA 26338
|||||
Db 45 CTGGCCAATATGGTGAACCCCTGCTCTACTAAAGATGTAATAA 2

RESULT 2
US-08-477-504A-65/c
; Sequence 65, Application US/08477504A
; Patent No. 5972353
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,504A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-477-504A-65

Query Match 0.3%; Score 76.8; DB 4; Length 105;
Best Local Similarity 83.7%; Pred. No. 1.le-06;
Matches 87; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 26235 ATCCAGCAGCTTTGGAGGCTGAGTGGTGAATCAGAGGTCAAGAGATCAAGACCATC 26294
|||||
Db 105 ATCCAGCAGCTTTGGAGGCGGAGGCTGGTGATCAAGGTCAAGGATTTGAGAGCAGC 46
|||||
QY 26295 CTGCCCAACATGGTGAACCCCGCTCTCTACTAAATAACAAAAA 26338
|||||
Db 45 CTGGCCAATATGGTGAACCCCTGCTCTACTAAAGATGTAATAA 2

RESULT 3
```

```
US-08-486-756A-65/c
; Sequence 65, Application US/08486756A
; Patent No. 5981711
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,756A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-486-756A-65

Query Match 0.3%; Score 76.8; DB 4; Length 105;
Best Local Similarity 83.7%; Pred. No. 1.le-06;
Matches 87; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 26235 ATCCAGCAGCTTTGGAGGCTGAGTGGTGAATCAGAGGTCAAGAGATCAAGACCATC 26294
|||||
Db 105 ATCCAGCAGCTTTGGAGGCGGAGGCTGGTGATCAAGGTCAAGGATTTGAGAGCAGC 46
|||||
QY 26295 CTGCCCAACATGGTGAACCCCGCTCTCTACTAAATAACAAAAA 26338
|||||
Db 45 CTGGCCAATATGGTGAACCCCTGCTCTACTAAAGATGTAATAA 2

RESULT 4
US-08-485-862B-65/c
; Sequence 65, Application US/08485862B
; Patent No. 5989838
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
```

```
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485.862B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/477,504
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-0727
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-485-862B-65

Query Match          0.3%; Score 76.8; DB 4; Length 105;
Best Local Similarity 83.7%; Pred. No. 1.1e-06;
Matches 87; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 26235 ATCCGAGCAGCTTTGGGAGCGCTGAGTGGTGAATCACGAGGTCAAGGATCAAGACCATC 26294
Db 105 ATCCGAGCAGCTTTGGGAGCGCGAGCTGCTGATCACAAGGTCAGGAGTTTGAGAGCAGC 46

Qy 26295 CTGGCCACATGGTGAAACCCGCTCTCTACTAAATACAAAAA 26338
Db 45 CTGGCCAAATATGGTGAAACCCGCTCTCTACTAAAGATGTAAAA 2

RESULT 5
US-08-787-739-65/c
; Sequence 65, Application US/08787739
; Patent No. 6027887
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 96
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 369 Pine Street, Suite 610
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/787,739
; FILING DATE: 24-JAN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,049
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/486,756
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/477,504
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/481,658
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,862
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,863
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 08/487,077
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-981-2034
; TELEFAX: 415-981-0332
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-787-739-65

Query Match          0.3%; Score 76.8; DB 5; Length 105;
Best Local Similarity 83.7%; Pred. No. 1.1e-06;
Matches 87; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 26235 ATCCGAGCAGCTTTGGGAGCGCTGAGTGGTGAATCACGAGGTCAAGGATCAAGACCATC 26294
Db 105 ATCCGAGCAGCTTTGGGAGCGCGAGCTGCTGATCACAAGGTCAGGAGTTTGAGAGCAGC 46

Qy 26295 CTGGCCACATGGTGAAACCCGCTCTCTACTAAATACAAAAA 26338
Db 45 CTGGCCAAATATGGTGAAACCCGCTCTCTACTAAAGATGTAAAA 2

RESULT 6
US-08-481-658B-65
; Sequence 65, Application US/08481658B
; Patent No. 5955075
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,658B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-481-658B-65

```

	Query Match	0.2%	Score 71.2;	DB 4;	Length 105;	
	Best Local Similarity	85.8%;	Pred. No. 1.4e-05;			
	Matches 91;	Conservative 0;	Mismatches 13;	Indels 2;	Gaps 1;	
QY 18903	TTTTTGTATTTT	TAGT	TAGAT	AGGGTTT	TCAAT	TGTCGCCAGCGTGCTCTCAAACTCC 18962
Db 2	TTTTTACATCTTT	TAGT	TAGAC	AGGGTTT	TCAAC	TATATGGCCAGCGTCTCTCAAACTCC 61
QY 18963	TGCCCTCAAGT	ATGCC	TCCTGCC	TCGGCC	TCCCAAT	TGCTGGGAT 19008
Db 62	TGACCT--	TGT	TGAT	TCCAC	CAGCGT	TCGGCCTCCCAAGTCTGGGAT 105

```

RESULT 7
US-08-477-504A-65
US-08-477-504A-65
Sequence 65, Application US/08477504A
Patent No. 597353
GENERAL INFORMATION:
APPLICANT: zavada, Jan
APPLICANT: Pastorekova, Silvia
APPLICANT: Pastorek, Jaromir
TITLE OF INVENTION: MN Gene and Protein
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leona L. Lauder
STREET: 6 Mariposa Court
CITY: Tiburon
STATE: California
COUNTRY: USA
ZIP: 94920
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,504A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/260,190
FILING DATE: 15-JUN-1994
ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-477-504A-65

Query Match 0.2%; Score 71.2; DB 4; Length 105;
Best Local Similarity 85.8%; Pred. No. 1.4e-05;
Matches 91; Conservative 0; Mismatches 13; Indels 2; Gaps 1;

Qy 18903 TTTTGTATTTTGTAGTAGAGATAGAGTTTCACATGCTGCCAGCGCTCTCAAACTCC 18962
Db 2 TTTTACATCTTTTGTAGTAGAGAGGTTTCACCATATATTTGGCAGGCTCTCTCAAACTCC 61

Qy 18963 TGCCTCAAGTGATCCTCCTCGCTGGCCTCCCAATGTGCTGGGAT 19008
Db 62 TGACCT--TGATGTCACACAGCCTTCGGCCTCCCAAAAGTCTGGGAT 105

RESULT 8
US-08-486-756A-65
; Sequence 65, Application US/08486756A
; Patent No. 5981711
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486.756A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-485-756A-65

Query Match 0.2%; Score 71.2; DB 4; Length 105;  
Best Local Similarity 85.8%; Pred. No. 1.4e-05;  
Matches 91; Conservative 0; Mismatches 13; Indels 2; Gaps 1;

QY 18903 TTTTGTATTTTAGTAGAGATAGGTTTTCACAAATGCTGGCCAGGCTGCTCTCAAACTCC 18962  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 2 TTTTACATCTTTTAGTAGAGACAGGTTTTCACCATATTGCCAGGCTGCTCTCAAACTCC 61

QY 18963 TGGCTCAAGTATCTCTCTGCTGGCTCCCAATGTGCTGGAT 19008  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 62 TGACCT--TGTGATCCAGCAGGCTCGGCTCCCAAGTGTGCTGGAT 105

RESULT 9  
US-08-485-862B-65  
Sequence 65, Application US/08485862B  
Patent No. 5989838  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485.862B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727

INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-485-862B-65

Query Match 0.2%; Score 71.2; DB 4; Length 105;  
Best Local Similarity 85.8%; Pred. No. 1.4e-05;  
Matches 91; Conservative 0; Mismatches 13; Indels 2; Gaps 1;

QY 18903 TTTTGTATTTTAGTAGAGATAGGTTTTCACAAATGCTGGCCAGGCTGCTCTCAAACTCC 18962  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 2 TTTTACATCTTTTAGTAGAGACAGGTTTTCACCATATTGCCAGGCTGCTCTCAAACTCC 61  
QY 18963 TGGCTCAAGTATCTCTCTGCTGGCTCCCAATGTGCTGGAT 19008  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 62 TGACCT--TGTGATCCAGCAGGCTCGGCTCCCAAGTGTGCTGGAT 105

RESULT 10  
US-08-787-739-65  
Sequence 65, Application US/08787739  
Patent No. 6027887  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 96  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 369 Pine Street, Suite 610  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/787,739  
FILING DATE: 24-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,049  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/486,756  
FILING DATE: 07-JUN-1995

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/481,658  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,862  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,863  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/487,077  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.4

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-981-2034  
TELEFAX: 415-981-0332  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-787-739-65

```

Query Match          0.2%; Score 71.2; DB 5; Length 105;
Best Local Similarity 85.8%; Pred. No. 1.4e-05;
Matches 91; Conservative 0; Mismatches 13; Indels 2; Gaps 1;

QY 18903 TTTTGTATTTTGTAGACATAGGGTTTCACAAATGCTGCCAGAGCTGCTCTCAAACTCC 18962
      ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Ddb 2 TTTTATACATCTTTAGTAGACAGAGGTTTCACCATATTTGCCAGAGCTGCTCTCAAACTCC 61
      ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 18963 TGCCTCAAGTGATCCCTCTGCTCGGCCCTCCCAATGCTGCTGGAT 19008
      ||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Ddb 62 TGAACCT--TGTGATCCACCAGCCTCGGCCCTCCCAAGAGTCTGGAT 105
      ||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
US-08-454-557C-91/c
; Sequence 91, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,557C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
US-08-454-557C-91

Query Match          0.2%; Score 65.4; DB 3; Length 84;
Best Local Similarity 86.7%; Pred. No. 0.00017;
Matches 72; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 26225 CACGCGCTGTAATCCACGACATTTGGGAGGCTGAGGTGGTAATCACGAGGTGAGAGAT 26284
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Ddb 83 CACGCTTGTAATCCACGACATTTGGGAGGCTGAGCGGGGATCATCAGAGTTCAGAGTT 24
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 26285 CAAGACCATCTCTGGCCAAATGG 26307
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Ddb 23 CGACACCAGCCTGATGAACATGG 1
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
US-08-340-426D-91/c
; Sequence 91, Application US/08340426D
; Patent No. 5948634

```



[illegible]

QY 7958 AAAAGAAACGATCAAGCCATGAAAACACATGAAGGAAA 7995  
 ||||| ||||| - || - ||||| - || - |||||  
 Db 61 AA 98

Search completed: June 20, 2000, 02:28:19  
Job time: 480829 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2000, 02:22:00 ; Search time 29135.9 seconds  
(without alignments)  
-968.286 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_168000\_197000  
Perfect score: 29001  
Sequence: 1 ATTTACAGATGGAGAAACCA.....GGATTAGGATCATGATCTC 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 370290

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_bal.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pli.\*
- 8: gb\_pl2.\*
- 9: gb\_pr1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_vl.\*
- 17: em\_fun.\*
- 18: em\_hum1.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_om.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pl.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_vl.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_ba1.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*
- 44: gb\_htg6.\*

- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*
- 50: gb\_pl3.\*
- 51: gb\_pr5.\*
- 52: gb\_htg8.\*
- 53: gb\_htg9.\*
- 54: gb\_htg10.\*
- 55: gb\_htg11.\*
- 56: gb\_htg12.\*
- 57: gb\_htg13.\*
- 58: gb\_htg14.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
c 1	96.8	0.3	108	10	HSIDLRLN2	X05250 Human LDL-r
c 2	92	0.3	108	10	HSIDLRLN2	X05250 Human LDL-r
c 3	86.2	0.3	108	10	HSIDLRLD1	X05249 Human LDL-r
c 4	86.2	0.3	108	10	HSIDLRLD2	X05251 Human LDL-r
c 5	83.4	0.3	107	9	HUMALCE162	M87924 Human carci
c 6	83	0.3	108	10	HSIDLRLD1	X05249 Human LDL-r
c 7	83	0.3	108	10	HSIDLRLD2	X05251 Human LDL-r
c 8	82.2	0.3	103	9	HUMALCE221	M87896 Human carci
c 9	81	0.3	108	11	HSU67803	U67803 Human small
c 10	81	0.3	108	11	HSU67808	U67808 Human small
c 11	78	0.3	107	9	HUMALCE162	M87924 Human carci
c 12	77.6	0.3	103	9	HUMALCE221	M87896 Human carci
c 13	76.6	0.3	108	10	HSIDLRLI12	X05248 Human LDL-r
c 14	76	0.3	103	13	HS8IC8R	X57789 Human sequ
c 15	75	0.3	110	11	HSU67807	U67807 Human small
c 16	74.4	0.3	104	9	HUMALCE272	M87899 Human carci
c 17	74.6	0.3	108	11	HSU67804	U67804 Human small
c 18	73.4	0.3	91	13	HUMUT8164A	L30244 Human STS U
c 19	73.6	0.3	97	9	HUMDLRLA2	M14180 Human low d
c 20	73.4	0.3	110	11	HSU67807	U67807 Human small
c 21	72.8	0.3	103	13	HS8IC8R	X57789 Human sequ
c 22	73	0.3	108	13	G43535	G43535 WIAF-2393-S
c 23	72.4	0.2	91	13	HUMUT8164A	L30244 Human STS U
c 24	72.4	0.2	110	9	HUMALCE43	M87900 Human carci
c 25	72	0.2	97	9	HUMDLRLA2	M14180 Human low d
c 26	72.2	0.2	108	9	HUMDLID03M5	D16965 Human HepG2
c 27	72	0.2	108	10	HSIDLRLI12	X05248 Human LDL-r
c 28	71.4	0.2	107	11	HSU67806	U67806 Human small
c 29	70.8	0.2	90	9	HUMDLRLFL	K03555 Human low d
c 30	70.8	0.2	100	13	HUMUT931A	L31299 Human STS U
c 31	69.4	0.2	108	13	G43535	G43535 WIAF-2393-S
c 32	68.6	0.2	84	5	AR051521	AR051521 Sequence
c 33	68.8	0.2	106	13	G32743	G32743 A009P31 Hum
c 34	68.4	0.2	95	13	HUMUT8002B	L30176 Human STS U
c 35	68.6	0.2	97	9	HUMDLRLD1	M14179 Human famli
c 36	67.8	0.2	108	9	HUMDLID03M5	D16965 Human HepG2
c 37	67.4	0.2	79	10	S73203	S73203 ALL-1 (tand
c 38	67.6	0.2	99	13	HUMUT7692A	L30306 Human STS U
c 39	67.8	0.2	108	3	AF185109S1	AF185109 Lasiiorhin
c 40	67.4	0.2	102	13	G32906	G32906 A009W09 Hum
c 41	67	0.2	97	9	HUMDLRLD1	M14179 Human famli
c 42	65.8	0.2	97	9	HUMDLRLA1	M14178 Human low d
c 43	65.2	0.2	95	10	HSSTHPKIB	X66361 H.sapiens m
c 44	65.4	0.2	100	13	HUMUT931A	L31299 Human STS U
c 45	65.6	0.2	106	13	G32743	G32743 A009P31 Hum

ALIGNMENTS

```

RESULT 1
HSLDLRN2/c
LOCUS      HSLDLRN2      108 bp      DNA      PRI      20-MAY-1992
DEFINITION Human LDL-receptor gene intron 14 fragment (normal gene).
ACCESSION  X05250
VERSION     X05250.1 GI:34337
KEYWORDS   Alu repetitive sequence; low density lipoprotein receptor.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 108)
AUTHORS    Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
            Williamson,R. and Humphries,S.
TITLE      Unequal crossing-over between two alu-repetitive DNA sequences in
            the low-density-lipoprotein-receptor gene. A possible mechanism for
            the defect in a patient with familial hypercholesterolaemia
            Eur. J. Biochem. 164 (1), 77-81 (1987)
JOURNAL    87161901
MEDLINE
COMMENT    See X05252 for deletion junction
            Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES   source
            Location/Qualifiers
            1..108
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            1..108
            /note="intron XIV fragment"
BASE COUNT 28 a 23 c 39 g 18 t
ORIGIN

Query Match      0.3%; Score 96.8; DB 10; Length 108;
Best Local Similarity 93.5%; Pred. No. 6.6e-07;
Matches 101; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 24901 CTTGGCTCAGTCAACCTCCGCTCTGGGTTCAAGCAATTCCTCAGCTCCGCTCCG 24960
|||
Db 108 CTTGGCTCAGTCAACCTCTGCTCTGGGTTCAAGCAATTCCTCAGCTCCGCTCCG 49

Qy 24961 AGTAGCTGGGATTACAGGCATGCCACCATGCTGGCTAATTTTGT 25008
|||
Db 48 AGTAGCTGGGATTACAGGCATGCCACCATGCTGGCTAATTTTGT 1

RESULT 2
HSLDLRN2
LOCUS      HSLDLRN2      108 bp      DNA      PRI      20-MAY-1992
DEFINITION Human LDL-receptor gene intron 14 fragment (normal gene).
ACCESSION  X05250
VERSION     X05250.1 GI:34337
KEYWORDS   Alu repetitive sequence; low density lipoprotein receptor.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 108)
AUTHORS    Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
            Williamson,R. and Humphries,S.
TITLE      Unequal crossing-over between two alu-repetitive DNA sequences in
            the low-density-lipoprotein-receptor gene. A possible mechanism for
            the defect in a patient with familial hypercholesterolaemia
            Eur. J. Biochem. 164 (1), 77-81 (1987)
JOURNAL    87161901
MEDLINE
COMMENT    See X05252 for deletion junction
            Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES   source
            Location/Qualifiers
            1..108
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            1..108
            /note="intron XIV fragment"
BASE COUNT 28 a 23 c 39 g 18 t
ORIGIN

```

## ORIGIN

```

Query Match      0.3%; Score 92; DB 10; Length 108;
Best Local Similarity 90.7%; Pred. No. 4.2e-06;
Matches 98; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 19072 ACAAATATACCCAGGCTGCTGCTAGTCCAGCTACTCGGGACACTGAG 19131
|||
Db 1 ACAAATATACCCAGGCTGCTGCTAGTCCAGCTACTCGGGACACTGAG 60

Qy 19132 GCAGGAGAATCTGAACCCAGGAGGAGGATTCAGTGCAGTGCAG 19179
|||
Db 61 GCAGGAGAATCTGAACCCAGGAGGAGGATTCAGTGCAGTGCAG 108

RESULT 3
HSLDLRD1
LOCUS      HSLDLRD1      108 bp      DNA      PRI      20-MAY-1992
DEFINITION Human LDL-receptor mutated gene with intron 12 deletion junction.
ACCESSION  X05249
VERSION     X05249.1 GI:34335
KEYWORDS   Alu repetitive sequence; low density lipoprotein receptor.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 108)
AUTHORS    Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
            Williamson,R. and Humphries,S.
TITLE      Unequal crossing-over between two alu-repetitive DNA sequences in
            the low-density-lipoprotein-receptor gene. A possible mechanism for
            the defect in a patient with familial hypercholesterolaemia
            Eur. J. Biochem. 164 (1), 77-81 (1987)
JOURNAL    87161901
MEDLINE
COMMENT    *source: hypercholesterol aemia
            See X05248 for corresponding normal gene sequence
            In the defective LDL-receptor gene the deletion occurred between two
            alu-repetitive sequences, that are in the same direction, the
            deletion eliminates exons 13 and 14 and changes the reading frame
            of the resulting spliced mRNA.
            Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES   source
            Location/Qualifiers
            1..108
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /cell_type="blood leukocytes from a patient with familial"
            misc_feature 1..108
            /note="deletion junction region intron 12/ intron 15"
BASE COUNT 20 a 40 c 20 g 28 t
ORIGIN

Query Match      0.3%; Score 86.2; DB 10; Length 108;
Best Local Similarity 87.9%; Pred. No. 4e-05;
Matches 94; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 24902 TTGGCTCACTCAACCTCCGCTCTGGGTTCAAGCAATTCCTCAGCTCCGCTCCGGA 24961
|||
Db 2 TCGCTCACTCAACCTCTGCTCTGGGTTCAAGCAATTCCTCAGCTCCGCTCCGGA 61

Qy 24962 GTAGCTGGGATTACAGGCATGCCACCATGCTGGCTAATTTTGT 25008
|||
Db 62 GTAGCTGGGATTACAGGCATGCCACCATGCTGGCTAATTTTGT 108

RESULT 4
HSLDLRD2/c
LOCUS      HSLDLRD2      108 bp      DNA      PRI      20-MAY-1992
DEFINITION Human LDL-receptor mutated gene with intron 14 deletion junction.
ACCESSION  X05251
VERSION     X05251.1 GI:34336
KEYWORDS   Alu repetitive sequence; low density lipoprotein receptor.

```

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 108)  
 AUTHORS Horsthemke, B., Beisiegel, U., Dunning, A., Havinga, J. R.,  
 Williamson, R. and Humphries, S.  
 TITLE Unequal crossing-over between two alu-repetitive DNA sequences in  
 the low-density-lipoprotein-receptor gene. A possible mechanism for  
 the defect in a patient with familial hypercholesterolaemia  
 Eur. J. Biochem. 164 (1), 77-81 (1987)  
 JOURNAL 87161901  
 MEDLINE  
 COMMENT \*source: hypercholesterol aemia  
 See X05250 for corresponding normal gene sequence  
 In the defective LDL-receptor gene the deletion occurred between two  
 alu-repetitive sequences, that are in the same direction, the  
 deletion eliminates exons 13 and 14 and changes the reading frame  
 of the resulting spliced mRNA.  
 Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.  
 FEATURES  
 source 1..108  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_type="blood leukocytes from a patient with familial"  
 /note="intron XIV fragment"  
 intron 1..108  
 BASE COUNT 28 a 20 c 40 g 20 t  
 ORIGIN  
 Query Match 0.3%; Score 86.2; DB 10; Length 108;  
 Best Local Similarity 87.9%; Pred. No. 4e-05;  
 Matches 94; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 QY 24902 TTGGCTCACTGAACCTCGCTCTCTGGTTCAAGCAATTCACGCTCCCGA 24961  
 Db 107 TCGCCTCACCAACCTCTGCTCTCTGGTTCAACCAATTTTCCTGCCCTCCCGA 48  
 QY 24962 GTAGCTGGGATTACAGGCACATGCCACCATGACTGGCTAATTTTGT 25008  
 Db 47 GTAGCTGGGATTACAGGCACCTGCCACCATGCTGGCTAATTTTGT 1  
 RESULT 5  
 HUMALCEL62 107 bp ss-RNA PRI 15-APR-1994  
 LOCUS Human carcinoma cell-derived Alu RNA transcript, clone CE162.  
 DEFINITION M87924  
 VERSION M87924.1 GI:174871  
 KEYWORDS Alu repeat.  
 SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 107)  
 AUTHORS Sinnett, D., Richer, C., Deragon, J. M. and Labuda, D.  
 TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of  
 post-transcriptional selection of master sequences  
 J. Mol. Biol. (1992) In press  
 JOURNAL  
 FEATURES  
 source 1..107  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_line="NTERA2D1"  
 /dev\_stage="embryo"  
 /sex="male"  
 /tissue\_type="carcinoma"  
 BASE COUNT 28 a 30 c 35 g 14 t  
 ORIGIN  
 Query Match 0.3%; Score 83.4; DB 9; Length 107;  
 Best Local Similarity 89.1%; Pred. No. 0.00012;

Matches 90; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
 QY 19131 GCAGAGAATACATTGAACCCAGGAGGAGAGATTGCAGTGCAGATCGCGCACT 19190  
 Db 5 GCAGAGAATGGCGTGAACCCGAGGAGGAGCTTGCAGTGCAGATCGCGCACT 64  
 QY 19191 GCATTCCAGCCCTGGGACAGAGCGAGACTCCATCTCAAAA 19231  
 Db 65 GCATCCAGCGCTGGGACAGAGCGAGACTCCGCTCAAAA 105  
 RESULT 6  
 HSLDLRD1/108 bp DNA PRI 20-MAY-1992  
 LOCUS Human LDL-receptor mutated gene with intron 12 deletion junction.  
 DEFINITION X05249  
 ACCESSION X05249  
 VERSION X05249.1 GI:34335  
 KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 108)  
 AUTHORS Horsthemke, B., Beisiegel, U., Dunning, A., Havinga, J. R.,  
 Williamson, R. and Humphries, S.  
 TITLE Unequal crossing-over between two alu-repetitive DNA sequences in  
 the low-density-lipoprotein-receptor gene. A possible mechanism for  
 the defect in a patient with familial hypercholesterolaemia  
 Eur. J. Biochem. 164 (1), 77-81 (1987)  
 JOURNAL 87161901  
 MEDLINE  
 COMMENT \*source: hypercholesterol aemia  
 See X05248 for corresponding normal gene sequence  
 In the defective LDL-receptor gene the deletion occurred between two  
 alu-repetitive sequences, that are in the same direction, the  
 deletion eliminates exons 13 and 14 and changes the reading frame  
 of the resulting spliced mRNA.  
 Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.  
 FEATURES  
 source 1..108  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_type="blood leukocytes from a patient with familial"  
 /note="deletion junction region intron 12/ intron 15"  
 misc\_feature 1..108  
 BASE COUNT 20 a 40 c 20 g 28 t  
 ORIGIN  
 Query Match 0.3%; Score 83; DB 10; Length 108;  
 Best Local Similarity 86.0%; Pred. No. 0.00014;  
 Matches 92; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
 QY 19072 ACAAAATTTAGCCAGCGCTGGTGGCATCTGCTGTAGTCCAGTACTCGGGACACTGAG 19131  
 Db 108 ACAAAATTTAGCCAGCGCTGGTGGCATCTGCTGTAGTCCAGTACTCGGGACACTGAG 49  
 QY 19132 GCAGGAGATCACTTGAACCCAGGAGCGAGATTTCAGTGAGCTCA 19178  
 Db 48 GCAGGAGAAATGTTTGAACCCAGGAGCGAGGTTGTGTGAGGCCGA 2  
 RESULT 7  
 HSLDLRD2 108 bp DNA PRI 20-MAY-1992  
 LOCUS Human LDL-receptor mutated gene with intron 14 deletion junction.  
 DEFINITION X05251  
 ACCESSION X05251  
 VERSION X05251.1 GI:34336  
 KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 108)

AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., HAVINGA,J.R.,  
Williamson,R. and Humphries,S.  
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in  
the low-density-lipoprotein-receptor gene. A possible mechanism for  
the defect in a patient with familial hypercholesterolaemia  
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)  
MEDLINE 87161901  
COMMENT \*source: hypercholesterol aemia  
See X05250 for corresponding normal gene sequence  
In the defective LDL-receptor gene the deletion occurred between two  
alu-repetitive sequences, that are in the same direction, the  
deletion eliminates exons 13 and 14 and changes the reading frame  
of the resulting spliced mRNA.  
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES  
source  
1. .108  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="blood leukocytes from a patient with familial"  
1. .108  
/note="intron XIV fragment"  
BASE COUNT 28 a 20 c 40 g 20 t  
ORIGIN

Query Match 0.3%; Score 83; DB 10; Length 108;  
Best Local Similarity 86.0%; Pred. No. 0.00014;  
Matches 92; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 19072 ACAAATAATAGCCAGGCTGTGGCATCTGCCTAGTCCAGTACTCGGGACACTGAG 19131  
Db 1 ACAAATAATAGCCAGGCTGTGGCATCTGCCTAGTCCAGTACTCGGGAGGCTGAG 60

Qy 19132 GCAGGAGATCACTGATACCCAGGAGGAGGATTCGAGTGCAGCTGA 19178  
Db 61 GCAGGAAATGGTTGATACCCAGGAGGAGGATTCGAGTGCAGCTGA 107

RESULT 8  
HUMALCE221 103 bp ss-RNA PRI 15-APR-1994  
LOCUS Human carcinoma cell-derived Alu RNA transcript, clone CE221.  
DEFINITION M87896  
ACCESSION M87896.1 GI:174874  
VERSION  
KEYWORDS Alu repeat.  
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 103)  
AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.  
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of  
post-transcriptional selection of master sequences  
JOURNAL J. Mol. Biol. (1992) In press  
FEATURES  
source  
1. .103  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="NTera2D1"  
/dev\_stage="embryo"  
/sex="male"  
/tissue\_type="carcinoma"  
BASE COUNT 25 a 27 c 33 g 18 t  
ORIGIN

Query Match 0.3%; Score 82.2; DB 9; Length 103;  
Best Local Similarity 87.4%; Pred. No. 0.00019;  
Matches 90; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 19097 ATCTGCCCTGTAGTCCCACTACTCGGGACACTGAGGCGAGGAGTAATCAATCCAGGA 19156  
Db 1 ATCTGCCCTGTAGTCCCACTACTACAGGGAAGCTAAGGCGAGGAGTAATCGCTTGAACCCCGGA 60

Qy 19157 GGCAGAGATTGCAGTGCAGCTGAGATCGCGCACATGCATTCACG 19199  
Db 61 GCGGAGGTTGCATGAGCCGAGATCGTGCATTCGACTCCAG 103

RESULT 9  
HSU67803 108 bp RNA PRI 01-AUG-1997  
LOCUS Human small cytoplasmic Alu transcript.  
DEFINITION U67803  
ACCESSION U67803  
VERSION U67803.1 GI:2289917  
KEYWORDS Alu.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 108)  
AUTHORS Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.  
TITLE cDNAs derived from primary and small cytoplasmic Alu (scAlu)  
transcripts  
JOURNAL J. Mol. Biol. 271 (2), 222-234 (1997)  
MEDLINE 97415756  
REFERENCE 2 (bases 1 to 108)  
AUTHORS Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.  
TITLE Direct Submission  
JOURNAL Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The  
Children's Hospital of Philadelphia, 1004F Abramson Research  
Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA

FEATURES  
source  
1. .108  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="tscAlu2"  
repeat\_region 1. .108  
/note="scAlu"  
/rpt\_family="Alu"  
/rpt\_type="dispersed"  
BASE COUNT 23 a 39 c 30 g 16 t  
ORIGIN

Query Match 0.3%; Score 81; DB 11; Length 108;  
Best Local Similarity 89.7%; Pred. No. 0.0003;  
Matches 87; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 18966 GCCTGTATCCAGCAGCTTTGGGAGGCCAAGCGGACAGATCAGAGTCAAGGATTGA 19025  
Db 1 GCCTGTATCCAGCAGCTTTGGGAGGCCGAGCGGGGATCAGAGTCAAGGATCGA 60

Qy 19026 GACCAGCTGACCAACATGGTGAACCCCTGTCTCTAC 19062  
Db 61 GACCATCTGCTACCAAGGTGAACCCCGTCTCTAC 97

RESULT 10  
HSU67808 108 bp RNA PRI 01-AUG-1997  
LOCUS Human small cytoplasmic Alu transcript.  
DEFINITION U67808  
ACCESSION U67808  
VERSION U67808.1 GI:2289922  
KEYWORDS Alu.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 108)  
AUTHORS Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.  
TITLE cDNAs derived from primary and small cytoplasmic Alu (scAlu)  
transcripts  
JOURNAL J. Mol. Biol. 271 (2), 222-234 (1997)  
MEDLINE 97415756  
REFERENCE 2 (bases 1 to 108)

AUTHORS Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.  
TITLE Direct Submission  
JOURNAL Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The Children's Hospital of Philadelphia, 1004F Abramson Research Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA

FEATURES  
source Location/Qualifiers  
1. .108  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Tscalu7"  
repeat\_region 1. .108  
/note="scAlu"  
/rpt\_family="Alu"  
/rpt\_type="dispersed"  
BASE COUNT 22 a 37 c 28 g 21 t  
ORIGIN

Query Match 0.3%; Score 81; DB 11; Length 108;  
Best Local Similarity 89.7%; Pred. No. 0.0003;  
Matches 87; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 19866 GCCTGTAATCCAGCACTTTGGAGGCCAAGCGGCAGATCAGGAGTCAGGAGTTGA 19025  
|||||  
Db 1 GCCTGTAATCCAGCACTTTGGAGGCCAAGCGGCAGATCAGGAGTTGA 60  
|||||

QY 19026 GACCAGCCTGACCAACATGCTGAACCCCTCTCTAC 19062  
|||||  
Db 61 GACCAGCCTGACCAACATGCTGAACCCCTCTCTTC 97  
|||||

RESULT 11  
HUMALCE162/c  
LOCUS HUMALCE162 107 bp ss-RNA PRI 15-APR-1994  
DEFINITION Human carcinoma cell-derived Alu RNA transcript, clone CE162.  
ACCESSION M87924  
VERSION M87924.1 GI:174871  
KEYWORDS Alu repeat.  
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 107)  
AUTHORS Sinnott,D., Richer,C., Deragon,J.-M. and Labuda,D.  
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of post-transcriptional selection of master sequences  
J. Mol. Biol. (1992) in press  
FEATURES Location/Qualifiers  
source 1. .107  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Ntera2D1"  
/dev\_stage="embryo"  
/sex="male"  
/tissue\_type="carcinoma"  
BASE COUNT 28 a 30 c 35 g 14 t  
ORIGIN

Query Match 0.3%; Score 78; DB 9; Length 107;  
Best Local Similarity 85.3%; Pred. No. 0.00094;  
Matches 87; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 20785 TTTTTCAGACAGAGTCTGACTCTGTGGCCAGGCTGGAGTGGCCACCATCTAGGCT 20844  
|||||  
Db 106 TTTTTCAGACAGAGTCTGCTGTGGCCAGGCTGGAGTGGCCACCATCTAGGCT 47  
|||||

QY 20845 CACTGAGCCTTACCTCTCGGGGTTCAAGCGATTCTCTGCC 20886  
|||||  
Db 46 CACTGCAAGCTCCGCTCCCGGGTTTCACGCCATTCTCTGCC 5  
|||||

RESULT 12

HUMALCE221/c  
LOCUS HUMALCE221 103 bp ss-RNA PRI 15-APR-1994  
DEFINITION Human carcinoma cell-derived Alu RNA transcript, clone CE221.  
ACCESSION M87896  
VERSION M87896.1 GI:174874  
KEYWORDS Alu repeat.  
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 103)  
AUTHORS Sinnott,D., Richer,C., Deragon,J.-M. and Labuda,D.  
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of post-transcriptional selection of master sequences  
J. Mol. Biol. (1992) in press  
FEATURES Location/Qualifiers  
source 1. .103  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Ntera2D1"  
/dev\_stage="embryo"  
/sex="male"  
/tissue\_type="carcinoma"  
BASE COUNT 25 a 27 c 33 g 18 t  
ORIGIN

Query Match 0.3%; Score 77.6; DB 9; Length 103;  
Best Local Similarity 86.0%; Pred. No. 0.0011;  
Matches 86; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 12799 CTGGAGTGCAGTGCAGCATCTCAGCTCACTGAACCTCCAATTCCTGAGTTCAAGCAT 12858  
|||||  
Db 103 CTGGAGTGCAGTGCAGCATCTCAGCTCACTGAACCTCCGCCCTCCCGGGTTCAAGCAT 44  
|||||

QY 12859 TCTCGTGCCTTCAGCTCCCAAGTAGCTGGGATTACAGGCA 12898  
|||||  
Db 43 TCTCGTGCCTTCAGCTCCCGGTAGCTGGGATTACAGGCA 4  
|||||

RESULT 13  
HSLDL112  
LOCUS HSLDL112 108 bp DNA PRI 20-MAY-1992  
DEFINITION Human LDL-receptor gene intron 12 fragment (normal gene) LDL - low density lipoprotein.  
ACCESSION X05248  
VERSION X05248.1 GI:34334  
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor; repetitive sequence.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 108)  
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R., Williamson,R. and Humphries,S.  
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in the low-density-lipoprotein-receptor gene. A possible mechanism for the defect in a patient with familial hypercholesterolaemia  
Eur. J. Biochem. 164 (1), 77-81 (1987)  
JOURNAL MEDLINE 87161901  
COMMENT see X05249 for deletion junction  
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES Location/Qualifiers  
source 1. .108  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
complement(<1..65)  
misc\_feature complement("Alu repeat")  
intron 1. .108  
/note="intron XII fragment"  
BASE COUNT 21 a 38 c 20 g 29 t  
ORIGIN

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 110)
SHAikh.T.H., Roy.A.M., Kim.J., Batzer.M.A. and Deininger.P.L.
cDNAs derived from primary and small cytoplasmic Alu (scAlu)
transcripts
J. Mol. Biol. 271 (2), 222-234 (1997)
97415756
2 (bases 1 to 110)
SHAikh.T.H., Kim.J., Batzer.M.A. and Deininger.P.L.
Direct Submission
Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The
Children's Hospital of Philadelphia, 1004F Abranson Research
Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA
FEATURES
            source
                1. .110
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="TscAlu6"
            repeat_region
                1. .110
                    /note="scAlu"
                    /rpt_family="Alu"
                    /rpt_type="dispersed"
BASE COUNT      26 a      39 c      24 g      21 t
ORIGIN
Query Match      0.3%; Score 75; DB 11; Length 110;
Best Local Similarity 84.8%; Pred. No. 0.003;
Matches 84; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 17338 GTAGAGATGGGGTTGCCCATGTTGGCCAGCGTGGTCTCGAATCTCTGGCGCTCAAGCGAT 17397
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 99 GGAAGATGGGGTTTCACCATGTTTGACAGCGTGGTCTTGAATCTTCGGGGCTCAAGTGAT 40
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 17398 CCACCTTGGCTTGGCGCTCCCAAAGTGGCTTAAGATTACAGGC 17436
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 39 CCACCCACTTTTGGCGCTCTCAAAGTGGCTGGGATTACAGGC 1

```

Search completed: June 20, 2000, 17:56:59  
Job time: 537503 sec

Query Match	0.3%	Score	76.6	DB	10	Length	108	
Best Local Similarity	82.2%	Pred. No.	0.0016					
Matches	88	Conservative	0	Mismatches	19	Indels	0	
							Gaps	0
QY	17548	TTGGTTCACTGCACACCTCTGCCCTCGGTTCAGCGCATTTCTTCGCTCAGTCTCCTGA	17607					
Db	2	TCGCCCTACCCACAACCTCTGCCCTCGGTTCAAACCATTTTCTTCGCTCAGCCTCCTTA	61					
QY	17608	GGAGCTGGGACTACAGACATGTGCACATACACCCAGCTAAATTTTGTGA	17654					
Db	62	GTAGCTGGGATTACAAGCATGTGCCACACCGCCGGCTGATTTTGTGA	108					
RESULT	14							
HS8IC8R		103 bp	DNA	STS			05-SEP-1991	
LOCUS								
DEFINITION								
ACCESSION								
VERSION								
KEYWORDS								
SOURCE								
ORGANISM								
REFERENCE								
AUTHORS								
TITLE								
JOURNAL								
REFERENCE								
AUTHORS								
TITLE								
JOURNAL								
REFERENCE								
AUTHORS								
TITLE								
JOURNAL								
MEDLINE								
COMMENT								
FEATURES								
source								

```
source
1. .103
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

BASE COUNT	29 a	28 c	23 g	22 t	1 others
ORIGIN					
Query Match	0.3%; Score 76; DB 13; Length 103;				
Best Local Similarity	84.2%; Pred. No. 0.0021;				
Matches 85; Conservative	0; Mismatches 16; Indels 0; Gaps 0;				
QY 17337	AGTAGAGATGGGGTTTCGCCCATGTTGGCCAGCGTGGTCTCGAACTCCTGGCCTCAAGCGA 17396				
Db	2 AGTAGAGATAGGGTTTCACCATGTTGGCCAGCGTGGTCAAGAACTCCTTGACCTTAAGTGA 61				
QY 17397	TCCACTTGGCCTTGGCCCTCCCAAGTGCCTAAGATTACAGGCA 17437				
Db	62 TCCACCCACCTCGACCTCCCAAGTGCNGAAATATTAGGCA 102				
RESULT 15					
HSU67807/c					
LOCUS	HSU67807	110 bp	RNA	PRI	01-AUG-1997
DEFINITION	human small cytoplasmic Alu transcript.				
ACCESSION	U67807				
VERSION	U67807.1 GI:2289921				
KEYWORDS	Alu.				
SOURCE	human.				
ORGANISM	Homo sapiens				

RESULT	15	HSU67807	110 bp	RNA	PRI	01-AUG-1997
LOCUS	HS067807/c	Human small cytoplasmic	Alu	transcript.		
DEFINITION	U67807	Human				
ACCESSION	U67807.1	GI:2289921				
VERSION	U67807.1	GI:2289921				
KEYWORDS	Alu.					
SOURCE	human.					
ORGANISM	Homo sapiens					

;







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2000, 05:58:17 ; Search time 953.1 Seconds  
(without alignments)  
7612.864 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_168000\_197000  
Perfect score: 29001  
Sequence: 1 ATTTACAGATGGAGAACCA.....GGATTAGGATCATGATCTC 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 433070

Minimum DB seq length: 10

Maximum DB seq length: 110

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	73.6	0.3	108	1 X12095	Human biallelic po
2	70	0.2	108	1 X12095	Human biallelic po
3	67.6	0.2	108	1 X12095	Human gene signatu
4	64	0.2	92	1 T65081	(dC-dA)n.(dG-dT)n
5	62.4	0.2	100	1 T24892	Human gene signatu
6	62.4	0.2	100	1 T24892	Human gene signatu
7	62.6	0.2	103	1 T26213	Human gene signatu
8	61.6	0.2	87	1 T21566	Human gene signatu
9	61.8	0.2	103	1 T20927	Human gene signatu
10	61.4	0.2	108	1 T26828	Human gene signatu
11	61	0.2	103	1 T26213	Human gene signatu
12	60.6	0.2	93	1 T22572	Human gene signatu
13	59.4	0.2	91	1 T25854	Human gene signatu
14	59.4	0.2	108	1 T26828	Human gene signatu
15	59	0.2	60	1 T65762	Repeat sequence fr
16	58	0.2	87	1 T21566	Human gene signatu
17	57.6	0.2	108	1 T25009	Human gene signatu
18	57	0.2	100	1 X12087	Human biallelic po
19	57	0.2	100	1 X12085	Human biallelic po
20	57	0.2	100	1 X12086	Human biallelic po
21	55.8	0.2	93	1 T24259	Human gene signatu
22	55.8	0.2	93	1 T24259	Human gene signatu
23	55.4	0.2	81	1 T24093	Human gene signatu
24	55.4	0.2	88	1 V39744	Microsatellite ana
25	55.2	0.2	95	1 T23131	Human gene signatu
26	55	0.2	99	1 T24420	Human gene signatu
27	54.6	0.2	91	1 T25854	Human gene signatu
28	53.6	0.2	69	1 Q29016	Probe to internal
29	53.6	0.2	91	1 T65740	Repeat sequence fr
30	53.2	0.2	73	1 Q34140	Sequence of a micr
31	53.4	0.2	99	1 T23728	Human gene signatu
32	53.4	0.2	100	1 X12087	Human biallelic po
33	53.4	0.2	100	1 X12085	Human biallelic po
34	53.4	0.2	100	1 X12086	Human biallelic po

c	35	53.6	0.2	103	1	T20927	Human gene signatu
c	36	52.8	0.2	93	1	T25688	Human gene signatu
c	37	53	0.2	97	1	T26728	Human gene signatu
c	38	52	0.2	69	1	Q29016	Probe to internal
c	39	52	0.2	86	1	Q34050	Microsatellite seq
c	40	51.6	0.2	93	1	T22572	Human gene signatu
c	41	51.2	0.2	56	1	T65707	Repeat sequence fr
c	42	51.2	0.2	99	1	T20931	Human gene signatu
c	43	50.8	0.2	97	1	T26728	Human gene signatu
c	44	50.8	0.2	100	1	T25604	Human gene signatu
c	45	50.6	0.2	110	1	T26288	Human gene signatu

ALIGNMENTS

RESULT	1
X12095	ID X12095 standard; DNA; 108 BP.
AC	X12095;
DT	30-MAR-1999 (first entry)
DE	Human biallelic polymorphic DNA fragment TIGR-A003M18a.
KW	Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW	detection; phenotypic typing; characteristic; infection; hereditary;
KW	autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW	treatment; marker; ss.
OS	Homo sapiens.
PN	WO9820165-A2.
PD	14-MAY-1998.
PF	05-NOV-1997; U20313.
PR	06-NOV-1996; US-030455.
PA	(WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI	Hudson T, Lander ES, Wang D;
DR	WPI; 98-286974/25.
PT	New isolated nucleic acid segments from the human genome - used for
PT	determining polymorphic forms for use in e.g. forensics, paternity
PT	testing or phenotypic typing for disease
PS	Claim 1; Page 219; 310pp; English.
CC	X12095-X12937 are human DNA fragments which contain biallelic polymorphic
CC	markers which have been isolated using the primers represented in
CC	X09121-X10268. The base occupying the polymorphic site is indicated by
CC	the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
CC	methods for determining polymorphic forms in an individual for use in
CC	e.g. forensics, paternity testing or for phenotypic typing for diseases
CC	such as acamaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC	muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC	hypercholesterolemia, polycystic kidney disease, hereditary
CC	spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC	haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC	syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC	autoimmune diseases, inflammation, cancer, diseases of the nervous
CC	system, infection by pathogenic microorganisms, and characteristics such
CC	as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC	endurance, fertility, and susceptibility or receptivity to particular
CC	drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC	segments can also be used to produce medicaments for the treatment or
CC	prophylaxis of such diseases.
CC	Sequence 108 BP; 19 A; 23 C; 28 G; 37 T;

Query Match	0.3%; Score 73.6; DB 1; Length 108;
Best Local Similarity	80.2%; Pred. No. 0.01;
Matches	85; Conservative 1; Mismatches 20; Indels 0; Gaps 0;
QY	10698 TTTTATAGAGATGGGTTTTCCTTAAACAGGCTGGTCTTGAACCTCTGACCT 10757
Db	3 TCTTTTCTAGAGATGAGGTTTTCCTTGGCCAGGATGCTCGAAGCTCTGACTT 62
QY	10758 CAAGTGATGCCACCTTGGCCCTCCATAGTCTGGGATTACAG 10803
Db	63 CAAGTGATCCGTGCTTGGCCCTCCAAAGTGCCTGGGATTATAG 108







```
RESULT 10
T26828/c
ID T26828 standard; cDNA to mRNA; 108 BP.
AC T26828:
DT 14-NOV-1996 (first entry)
DE Human gene signature HUMGS09078.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
human; cloning; mapping; non-biased library; diagnosis; detection;
cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
WPI; 95-206931/27.
DR Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 2182; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 108 BP; 18 A; 33 C; 23 G; 28 T;

Query Match 0.2%; Score 61.4; DB 1; Length 108;
Best Local Similarity 79.8%; Pred. No. 0.57;
Matches 71; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 18934 ATCATATACTAGGCTGGCGGTGATGGCTCAGCCCTGTATCCACGACTTTGGGAGGCC 18993
DB 90 AAGAAATAAACAGCGCGGCGTGGCTCATGCCCTGTAAACCCAGCACTATGGGAGGCC 31

QY 18994 AAGCGGACAGATCAGGAGTTCAGGAGTT 19022
DB 30 GANACGGCGGATGACGAGTTCAGGAGAT 2

RESULT 11
T26213/c
ID T26213 standard; cDNA to mRNA; 103 BP.
AC T26213:
DT 13-NOV-1996 (first entry)
DE Human gene signature HUMGS08452.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
human; cloning; mapping; non-biased library; diagnosis; detection;
cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
WPI; 95-206931/27.
```

```
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 2029; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 103 BP; 33 A; 21 C; 25 G; 23 T;

Query Match 0.2%; Score 61; DB 1; Length 103;
Best Local Similarity 75.2%; Pred. No. 0.65;
Matches 76; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 27386 TTTTGTGTTTATAGACACGATCTTATCTGTACCCAGGCTGTAGCTAGTGGCCAA 27445
DB 102 TTTTGTGTTTATAGACATGTTCTTACTTGTGGCCAGGCTGGAGTGGGTGCCA 43

QY 27446 TCATTGCTCACTCAGCCTCAACTCTCTGGGCTCCAGTAAT 27486
DB 42 TCATAGCTCAGCTATACACCAAACTCTGGGCTCAAGTGAT 2

RESULT 12
T22572/c
ID T22572 standard; cDNA to mRNA; 93 BP.
AC T22572:
DT 01-OCT-1996 (first entry)
DE Human gene signature HUMGS04188.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
human; cloning; mapping; non-biased library; diagnosis; detection;
cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
WPI; 95-206931/27.
DR Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1159; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
```









GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2000, 21:22:20 ; Search time 13789.4 seconds  
(without alignments)  
8524.491 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_168000\_197000  
Perfect score: 29001  
Sequence: 1 ATTACAGATGGAGAAACCA.....GGATTAGGATCATGATCTC 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 156056

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	89	0.3	105	105	AQ282107	RPC11-94
2	88.8	0.3	108	84	B65160	CIT-HSP-201
3	87.2	0.3	109	94	AQ028426	CIT-HSP-2
4	86.8	0.3	106	97	AQ0703692	ag81a10.r
5	85	0.3	101	35	AA583697	nn58f10.s
6	85.2	0.3	106	63	AI991750	wt48e01.x
7	84.8	0.3	107	33	AA385808	EST99495
8	85	0.3	109	22	H11143	ym09c06.r1
9	85.2	0.3	109	24	N25299	yw52c09.s1
10	85	0.3	110	30	AA244245	nc07a04.s
11	85	0.3	110	30	AA244245	nc07a04.s
12	84.6	0.3	107	35	AA565533	nk42b11.s
13	83.8	0.3	103	108	AQ535244	RPC1-11-3
14	84	0.3	109	103	AQ200347	RPC11-43
15	83.2	0.3	106	94	AQ046231	RPC111-36
16	82.8	0.3	102	36	AA654562	nt75f10.s
17	82.8	0.3	106	30	AA250812	zs06a05.s
18	82.8	0.3	107	24	H67040	yu68c01.r1
19	82.4	0.3	109	30	AA243009	zt25h02.s
20	81.8	0.3	103	108	AQ535244	RPC1-11-3
21	82	0.3	106	105	AQ282340	RPC111-80
22	81.6	0.3	104	29	AA129957	zn86h04.r
23	81.6	0.3	105	30	AA218889	zq15d04.s
24	81.8	0.3	109	84	B17434	345K2.TVB C
25	81.8	0.3	109	94	AQ028426	CIT-HSP-2
26	81.4	0.3	108	84	B65160	CIT-HSP-201
27	81	0.3	105	28	AA078003	7H12D08 C
28	81	0.3	106	44	AI249096	qn73g09.x
29	80.6	0.3	103	108	AQ584425	RPC1-11-4
30	80.6	0.3	104	105	AQ321855	RPC111-11
31	80.6	0.3	106	94	AQ062963	CIT-HSP-2
32	80.8	0.3	108	84	B15423	345B10.TV C
33	80.2	0.3	101	39	AA835205	ak64h01.s
34	80.4	0.3	106	38	AA812141	0D48h02.s
35	80.4	0.3	106	106	AQ14071	RPC1-11-1
36	79.8	0.3	101	33	AA381369	EST94442
37	80	0.3	104	108	AQ544583	CITBI-El-
38	79.6	0.3	102	84	B48088	RPC111-4N6
39	79.8	0.3	107	62	AI933497	wn74d02.x
40	79.8	0.3	110	39	AA897366	am06h02.s
41	79.4	0.3	105	105	AQ276193	CITBI-El-
42	79.4	0.3	107	24	H67040	yu68c01.r1
43	79.6	0.3	110	106	AQ386882	RPC111-13
44	79.6	0.3	110	109	AQ634950	RPC1-11-4
45	79.2	0.3	100	30	AA252633	zq43g05.r

## ALIGNMENTS

RESULT 1  
 AQ282107 105 bp DNA 27-APR-1999  
 LOCUS RPC111-94B21.TJ RPC1-11 Homo sapiens genomic clone RPC1-11-94B21,  
 DEFINITION genomic survey sequence.

ACCESSION AQ282107 GI:3907976  
 VERSION AQ282107  
 KEYWORDS GSS.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 105)  
 AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,  
 Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.  
 Title Use of human BAC End Sequences for Sequence-Ready Map Building

JOURNAL Unpublished (1998)  
 COMMENT Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadamst@igrr.org  
 Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pieter de Jong (pieter@igrr.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/humgen/bac\_end\_search/bac\_end\_search.html  
 Seq primer: SP6  
 Class: BAC ends.

## FEATURES

## source

1..105  
 /organism="Homo sapiens"  
 /db\_xref="GDB:7535756"  
 /db\_xref="taxon:9606"  
 /clone="RPC1-11-94B21"  
 /clone\_lib="RPC1-11"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /notes="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
 RPC111 Human Male BAC Library"  
 BASE COUNT 26 a 31 c 30 g 18 t  
 ORIGIN

Query Match 0.3%; Score 89; DB 105; Length 105;  
 Best Local Similarity 90.5%; Pred. No. 0.34; Indels 0; Gaps 0;  
 Matches 95; Conservative 0; Mismatches 10;

QY 18960 GCTCACGCTGTATCCAGCACATTTGGGAGGCCAGCGGACAGATCAGAGTTCAGGA 19019  
 |||||  
 Db 1 GCTCACGCTGTATCCAGCACATCTGGGAGGCCAGCGGATGATCAGAGGCGCATGA 60  
 QY 19020 GTTTGACACGAGCTGACCAACATGTTGAACCTGTCTCTACTA 19064  
 |||||  
 Db 61 GTACGAGACCGAGCTGACCAACATGTTGAACCTGTCTCTACTA 105

## RESULT 2

B65160/c 2  
 LOCUS B65160 108 bp DNA GSS 21-JUN-1998  
 DEFINITION CIT-HSP-2017G2.TRB CIT-HSP Homo sapiens genomic clone 2017G2,  
 genomic survey sequence.  
 B65160  
 ACCESSION B65160.1 GI:2639138  
 VERSION B65160.1  
 KEYWORDS GSS.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 108)  
 AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,  
 Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,  
 Simon,M. and Venter,J.C.  
 Title Use of a random BAC End Sequence Database for Sequence-Ready Map Building

JOURNAL Unpublished (1997)  
 COMMENT Other\_GSSs: CIT-HSP-2017G2.TPB  
 Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadamst@igrr.org  
 Clones are available from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/humgen/bac\_end\_search/bac\_end\_search.html  
 Seq primer: M13 Reverse  
 Class: BAC ends.

## FEATURES

## source

1..108

```

/organism="Homo sapiens"
/db_xref="GDB:7043860"
/db_xref="taxon:9606"
/clone="201762"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII"
HindIII"
BASE COUNT      26 a   27 c   34 g   21 t
ORIGIN

Query Match      0.3%; Score 88.8; DB 84; Length 108;
Best Local Similarity 88.9%; Pred. No. 0.36;
Matches 96; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 24883 AGAATGCGGTGTGTGCTTGGCTCACTGCAACCTCCGCTCTGGGTCAAGCAATTC 24942
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 108 AGTGTGCAGTGTGTATGATCTTGGCTCACTGCAACCTCCGCTCTGGGTCAAGCAATTC 49

QY 24943 CCATGCTCAGCTCCGCTGAGTGGGATTACAGGCACATGCCACCA 24990
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 48 TCCTGCTCAGCTCTCTGAGTAGCTGGGATTACAGGCACATGCCACCA 1

RESULT 3
LOCUS      AQ028426      109 bp      DNA      GSS      30-JUN-1998
DEFINITION CIT-HSP-2313G15.TF CIT-HSP Homo sapiens genomic clone 2313G15,
genomic survey sequence.
ACCESSION      AQ028426
VERSION      AQ028426.1 GI:3268648
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 109)
AUTHORS      Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
TITLE      Use of a random BAC End Sequence Database for Sequence-Ready Map
BUILDING      Building (1998)
JOURNAL      Unpublished (1998)
COMMENT      Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES             Location/Qualifiers
     source           1..109
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2313G15"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT      19 a   36 c   25 g   29 t
ORIGIN

Query Match      0.3%; Score 87.2; DB 94; Length 109;
Best Local Similarity 88.0%; Pred. No. 0.54;

```

```

Matches 95; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 24586 TTTTTCCTGAATGAGTCTCACTCTGTGCCAGCTGGAGTACAGTGGCACAATCTTG 24645
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2 TGTTCCTGAGACGACCTCACTCTGTCCACCAGCTGGAGTGCAGTGGCAGACTCTGA 61

QY 24646 GTTACTGCAACTCCACTCTCTGGTTCACGCGAGTCTCTCTGACTCA 24693
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 62 GCTCACTCAACTCCACTCTCTGGTTCACGCGATTCTCTGCTGCCTCA 109

RESULT 4
LOCUS      AA703692/c      106 bp      mRNA      EST      24-DEC-1997
DEFINITION ag1a10.r1 Stratagene hMT neuron (#937233) Homo sapiens cDNA clone
IMAGE:1140858 5' similar to contains Alu repetitive element.; mRNA
sequence.
ACCESSION      AA703692
VERSION      AA703692.1 GI:2713610
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 106)
AUTHORS      Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
Theising,B., White,Y., Wyllie,T., Waterston,K. and Wilson,R.
TITLE      WashU-NCI human EST project
JOURNAL      Unpublished (1997)
COMMENT      On Sep 12, 1996 this sequence version replaced gi:1397630.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LUNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 53.
FEATURES             Location/Qualifiers
     source           1..106
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1140858"
                     /clone_lib="Stratagene hMT neuron (#937233)"
                     /dev_stage="hMT neurons"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally. Primer: Oligo dT.
                     Differentiated, post mitotic hMT neurons. Average insert
                     size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
                     GAATTCGCCAGCAG 3' -3' adaptor sequence: 5'
                     CTCGAGTTTCTTTTCTTTT 3'"
BASE COUNT      19 a   29 c   29 g   29 t
ORIGIN

Query Match      0.3%; Score 86.8; DB 37; Length 106;
Best Local Similarity 88.7%; Pred. No. 0.61;
Matches 94; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 18963 CAGGCTCTAATCCAGCAGCTTTGGGAGCCGAGCGGACATCAGAGGTCAGAGTT 19022
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 106 CAGGCTCTAATCCAGCAGCCTTTGGGAGGCTGAGCGGCGCATCATCAGAGTCAGAGAT 47

QY 19023 TGAGACCAGCCTTGACCAACATGTTGTAACCTCTCTCTACTAACAA 19068
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 46 CGAGACCATCTCTGGCTTAACACGGTGAACCTCTCTCTACTAAAAA 1

```



Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L., Wei, X.F., Wing, J., Xu, C., Yu, G.D., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.

**TITLE**  
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence

**JOURNAL**  
Nature 377 (6547 Suppl), 3-174 (1995)

**MEDLINE**  
12140200

**COMMENT**  
On Jan 25, 1995 this sequence version replaced gi:637865.

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tldb/hgi/hgi.html>)

Seq primer: M13 Reverse.

**FEATURES**

Location/Qualifiers

1..107

/organism="Homo sapiens"

/db\_xref="ATCC (inhost):189984"

/db\_xref="taxon:9606"

/clone\_lib="Thyroid"

/dev\_stage="adult"

/note="organ: thyroid gland; Vector: pBluescript SK-;

Site\_1: ECORI; Site\_2: XhoI"

**BASE COUNT** 16 a 34 c 28 g 26 t 3 others

**ORIGIN**

**Query Match**  
Best Local Similarity 0.3%; Score 84.8; DB 33; Length 107;  
Matches 92; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

**QY 19121** GGGACACTGAGCAGAGAGATCACTTGACCCAGGAGGAGAGATTCAGTGCAGCTGAGA 19180

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

**Db 107** GGGAGCTGAGCAGGAGAACCGCTGTAACCCAGGAGGAGCTTGCAGTGCAGCTGAGA 48

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

**QY 19181** TCGCGCACTGCATCCACGCTGGGAGACAGAGCAGACTCCATCTC 19227

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

**Db 47** TCTCGCCACTGCATCCAACTCGGGTACAGGAGCAGACTGTCTC 1

**RESULT 8**

**H11143/c**

**LOCUS** H11143 109 bp mRNA EST 26-JUN-1995

**DEFINITION** ym09c06.r1 Soares infant brain lntb Homo sapiens cDNA clone

IMAGE:47310 5' similar to contains Alu repetitive element; contains

MER22 repetitive element ;, mRNA sequence.

**ACCESSION** H11143

**VERSION** H11143.1 GI:875963

**KEYWORDS** EST.

**SOURCE** human.

**ORGANISM** Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE** 1 (bases 1 to 109)

**AUTHORS** Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,

Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and

Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

**TITLE** On May 5, 1995 this sequence version replaced gi:798506.

**JOURNAL** Contact: Wilson RK

**COMMENT** Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: estewatson.wustl.edu

Insert Size: 1316

Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the

IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.

Putative full length read

Insert Length: 1316 Std Error: 0.00

Seq primer: M13RP1

High quality sequence stop: 363.

**FEATURES**

Location/Qualifiers

1..109

/organism="Homo sapiens"

/db\_xref="GDB:419851"

/db\_xref="taxon:9606"

/clone="IMAGE:47310"

/clone\_lib="Soares infant brain lntb"

/sex="female"

/dev\_stage="73 days post natal"

/lab\_host="DH10B (ampicillin resistant)"

/note="Organ: whole brain; Vector: Lfamid BA; Site\_1: Not

I; Site\_2: Hind III; 1st strand cDNA was primed with a Not

I - oligo(dT) primer [5'

AACTGGAGAATTCGCGCGCAGCAATTTTTTTTTTTT 3'];

double-stranded cDNA was ligated to Hind III adaptors

(Pharmacia), digested with Not I and directionally cloned

into the Not I and Hind III sites of the Lfamid BA vector.

Library went through one round of normalization. Library

constructed by Bento Soares and M.Fatima Bonaldo."

**BASE COUNT** 27 a 28 c 36 g 18 t

**ORIGIN**

**Query Match**  
Best Local Similarity 0.3%; Score 85; DB 22; Length 109;  
Matches 94; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

**QY 2499** TGAGACGGAGTCTCACTCTGTACCCAGGCTGGAGTGCAGTGCAGTCTCGGCTCACT 2558

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

**Db 109** TGAGAAGGCGTCTCACTCTGTACCCAGGCTGGAGTGCAGTGCAGTCTCACTCACT 50

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

**QY 2559** GCACCTCCGCGCTCCCGGTTCAAGCGATTCTCTCGCTCAGGCTCCCGA 2607

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

**Db 49** GGAACCTCGCTCCCGGTTCAAGGATTCTCTCTGCTTAGCCTCTCTGA 1

**RESULT 9**

**N25299/c**

**LOCUS** N25299 109 bp mRNA EST 28-DEC-1995

**DEFINITION** yw52c09.s1 Weizmann Olfactory Epithelium Homo sapiens cDNA clone

IMAGE:255856 3' similar to contains Alu repetitive element;; mRNA

sequence.

**ACCESSION** N25299

**VERSION** N25299.1 GI:1139449

**KEYWORDS** EST.

**SOURCE** human.

**ORGANISM** Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE** 1 (bases 1 to 109)

**AUTHORS** Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,

Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,

Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,

Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,

Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,

Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.

and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags

Genome Res. 6 (9), 807-828 (1996)

**JOURNAL** 97044478

**MEDLINE**

**COMMENT** On Apr 14, 1993 this sequence version replaced gi:837394.

Seq primer: -41m13 fwd. ET from Amersham



High quality sequence stop: 90.

# FEATURES

source

1. .110  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1007406"  
/clone\_lib="NCI\_CGAP\_Prl"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"

/note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st strand cDNA was primed with oligo(dT)17 on 50 ng of DNase-treated, total cellular RNA obtained from 5,000-10,000 microdissected, histologically normal prostate epithelial cells. Double-stranded cDNA was ligated to EcoRI adaptors, 5 cycles of PCR applied to the cDNA with an adaptor-specific primer, and the resulting PCR product subcloned into pAMP10 by the UDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman."

BASE COUNT 17 a 26 c 28 g 38 t 1 others

ORIGIN

Query Match 0.3%; Score 85; DB 30; Length 110;  
Best Local Similarity 85.5%; Pred. No. 0.95;  
Matches 94; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 19128 TGAGGCGAGAGATCACTTGAACCCAGGAGGAGAGATTGCGAGCTGAGATCGCGCC 19187  
Db 110 TGAGGCGAGAGATCTTTGAACCCAGGAGGAGAGATTGCGAGCTGAGATCGCGCC 51  
QY 19188 ACTGATTCAGCGCTGGGAGACAGCGAGAGATCCATCTCAAAATTA 19237  
Db 50 ACTGATTCAGCGCTGGGAGACAGATCAAGACTCCATCTCAAAATTA 1

RESULT 12

AA565533/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)

On Sep 12, 1996 this sequence version replaced gi:1393355.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1350

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Stratagene, Inc., David B. Krizman,

Ph.D.

cDNA Library Arraying: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone Distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html

Insert Length: 1661 Std Error: 0.00

Seq primer: -40m13 fwd. Et from Amerham

High quality sequence stop: 87.

Location/Qualifiers

FEATURES

source

1. .107  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1016157"  
/clone\_lib="NCI\_CGAP\_GC2"  
/tissue\_type="germ cell tumor"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Bulk germ cell tumor. 5' adaptor sequence: 5' GAATTCGGCAGAG 3' 3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."

BASE COUNT 22 a 34 c 26 g 25 t

ORIGIN

Query Match 0.3%; Score 84.6; DB 35; Length 107;  
Best Local Similarity 86.9%; Pred. No. 1.1;  
Matches 93; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 19093 TGGCATCTGCCTTACTCCAGCTACTCGGACACTGAGCAGGAGAACTTGAACCC 19152  
Db 107 TGGTGTGCTGTAATCCAGCTACTCAGGAGGCTGAGCAGCACTTGAACCT 48

QY 19153 AGGAGCGAGAGATTGCGAGCTGAGATCGCGCCACTGCATTCCAG 19199

Db 47 GGGAGCGAGAGCTTGCAGTGCAGCTGAGATTGAGCCACTGCATCCAG 1

RESULT 13

AA535244

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbe@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from

Research Genet cs (info@resgen.com). BAC end search page:

http://www.tigr.org/tldb/hungen/bac\_end\_search/bac\_end\_search.html.

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1. .103

/organism="Homo sapiens"

/db\_xref="GDB:762153"

/db\_xref="taxon:9606"

/clone="RPCI-11-317H22"

/clone\_lib="RPCI-11"

/sex="Male"

/cell\_type="Lymphocytes"

/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;

RPCI11 Human Male BAC Library"

```
BASE COUNT      31 a      27 c      27 g      18 t
ORIGIN

Query Match      0.3%; Score 83.8; DB 108; Length 103;
Best Local Similarity 88.3%; Pred. No. 1.3;
Matches 91; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 19961 CCAGCATTGTTGGAGCGCGAGCGTGCAGATCACATTGAGGACAGGAGTTCAAGACCAAGCC 20020
|||||
Db 1 CCAGCATTGTTGGAGCGCGAGCGCGAGCGTGCAGATCACATTGAGGACAGGAGTTCAAGACCAAGCC 60
|||||

Qy 20021 TGGCTAACATGCGCAAAACCCCATCTCTACTATAAAATACAAAA 20063
|||||
Db 61 TGGCCAAACATGTTGAAACCCCGCTCTCTCTATATAATACAAAA 103
|||||

RESULT 14
LOCUS      AQ200347
DEFINITION RPC111-43B21..TJ RPC1-11 Homo sapiens genomic clone RPC1-11-43B21,
genomic survey sequence.
ACCESSION      AQ200347
VERSION      AQ200347.1 GI:3612546
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 109)
AUTHORS      Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE      Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL      Unpublished (1998)
COMMENT      Other_GSSs: RPC111-43B21.TK
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPC1-11. For BAC
library availability, please contact Pieter de Jong
(pleter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Class: BAC ends.

FEATURES             Location/Qualifiers
     source           1..109
                     /organism="Homo sapiens"
                     /db_xref="GDB:7516172"
                     /db_xref="taxon:9606"
                     /clone="RPC1-11-43B21"
                     /clone_lib="RPC1-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"

BASE COUNT      27 a      31 c      29 g      22 t
ORIGIN

Query Match      0.3%; Score 84; DB 103; Length 109;
Best Local Similarity 86.1%; Pred. No. 1.2;
Matches 93; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 20012 AGACCAAGCTGGCTAACATGGCAAAACCCCATCTCTACTATAAAATACAAAATTAACCAAG 20071
|||||
Db 2 ATACCAAGCTGGCGACACGCGTGAACCCCATCTCTACTATAAAATACAAAATTAATAGCCAG 61
|||||

Qy 20072 GCGTGGTGGTGACGCCCTGTATATCCAGCTACTCTGGAGGCTGAGGCA 20119
|||||
```

```
Db 62 CTGTGGTGGCGCAGCGCTGTGTATCCAGCTACTCGGGAGGCTGAGCA 109
|||||

RESULT 15
LOCUS      AQ046231/c
DEFINITION RPC111-36G16..TJ RPC1-11 Homo sapiens genomic clone RPC1-11-36G16,
genomic survey sequence.
ACCESSION      AQ046231
VERSION      AQ046231.1 GI:3315158
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 106)
AUTHORS      Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
Venter,J.C.
TITLE      Use of BAC End Sequences for Sequence-Ready Map Building (1998)
JOURNAL      Unpublished (1998)
COMMENT      Other_GSSs: RPC111-36G16.TK
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPC1-11. For BAC
library availability, please contact Pieter de Jong
(pleter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Class: BAC ends.

FEATURES             Location/Qualifiers
     source           1..106
                     /organism="Homo sapiens"
                     /db_xref="GDB:7513599"
                     /db_xref="taxon:9606"
                     /clone="RPC1-11-36G16"
                     /clone_lib="RPC1-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"

BASE COUNT      37 a      21 c      28 g      20 t
ORIGIN

Query Match      0.3%; Score 83.2; DB 94; Length 106;
Best Local Similarity 87.5%; Pred. No. 1.5;
Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 27795 CTCTGCCCTCCTGTGTTTAAAGCGATTCTCTGCTCAGACTCCTGAGTAGCTGAGATTACA 27854
|||||
Db 105 CTCCTCCTACCGTGTGTTTAGCGATTCTCATGCCAGCCTCTTGAGTAGTGGGACTATA 46
|||||

Qy 27855 GGTGCCTGCCACCATGCCAGCTAATTTTATATTTTAGTAGA 27898
|||||
Db 45 GGTGTCTGCCACCATGCCAGCTAATTTTGTATTTTAGTAGA 2
|||||

Search completed: June 20, 2000, 09:51:12
Job time: 508439 sec
```

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2000, 02:28:19 ; Search time 599.42 seconds  
(without alignments)  
6288.907 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_168000\_197000

Perfect score: 23001

Sequence: 1 ATTTACAGATGGAGAACCA.....GGATTAGGATCATGATCTC 29001

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 374504

Minimum DB seq length: 10

Maximum DB seq length: 110

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	82.6	0.3	105	4	US-08-481-658B-65
C 2	82.6	0.3	105	4	US-08-477-504A-65
C 3	82.6	0.3	105	4	US-08-486-756A-65
C 4	82.6	0.3	105	4	US-08-485-862B-65
C 5	82.6	0.3	105	5	US-08-787-739-65
C 6	68.6	0.2	84	3	US-08-454-557C-91
C 7	68.6	0.2	84	4	US-08-340-426D-91
C 8	68.6	0.2	84	4	US-08-450-673C-91
C 9	68.6	0.2	84	6	PCT-US95-17111A-91
C 10	65.6	0.2	105	4	US-08-481-658B-65
C 11	65.6	0.2	105	4	US-08-477-504A-65
C 12	65.6	0.2	105	4	US-08-486-756A-65
C 13	65.6	0.2	105	4	US-08-485-862B-65
C 14	65.6	0.2	105	5	US-08-787-739-65
C 15	64	0.2	92	1	US-08-222-177A-430
C 16	59.4	0.2	78	3	US-08-454-557C-70
C 17	59.4	0.2	78	4	US-08-340-426D-70
C 18	59.4	0.2	78	4	US-08-450-673C-70
C 19	59.4	0.2	78	6	PCT-US95-17111A-70
C 20	59	0.2	60	1	US-08-222-177A-244
C 21	57.8	0.2	78	3	US-08-454-557C-70
C 22	57.8	0.2	78	4	US-08-340-426D-70
C 23	57.8	0.2	78	4	US-08-450-673C-70
C 24	57.8	0.2	78	6	PCT-US95-17111A-70
C 25	56	0.2	85	3	US-08-454-557C-92
C 26	56	0.2	85	4	US-08-340-426D-92
C 27	56	0.2	85	4	US-08-450-673C-92

C	28	56	0.2	85	6	PCT-US95-17111A-92	Sequence 92, Appl
	29	55	0.2	84	3	US-08-454-557C-91	Sequence 91, Appl
	30	55	0.2	84	4	US-08-340-426D-91	Sequence 91, Appl
	31	55	0.2	84	4	US-08-450-673C-91	Sequence 91, Appl
	32	55	0.2	84	6	PCT-US95-17111A-91	Sequence 166, Appl
	33	53.6	0.2	91	1	US-08-222-177A-166	Sequence 65, Appl
	34	51.2	0.2	56	1	US-08-222-177A-65	Sequence 60, Appl
C	35	50.4	0.2	60	3	US-08-454-557C-60	Sequence 60, Appl
C	36	50.4	0.2	60	4	US-08-340-426D-60	Sequence 60, Appl
C	37	50.4	0.2	60	4	US-08-450-673C-60	Sequence 60, Appl
C	38	50.4	0.2	60	6	PCT-US95-17111A-60	Sequence 60, Appl
	39	50	0.2	76	3	US-08-454-557C-69	Sequence 69, Appl
	40	50	0.2	76	4	US-08-340-426D-69	Sequence 69, Appl
	41	50	0.2	76	4	US-08-450-673C-69	Sequence 69, Appl
	42	50	0.2	76	6	PCT-US95-17111A-69	Sequence 69, Appl
	43	50	0.2	83	4	US-08-481-658B-66	Sequence 66, Appl
	44	50	0.2	83	4	US-08-477-504A-66	Sequence 66, Appl
	45	50	0.2	83	4	US-08-486-756A-66	Sequence 66, Appl

ALIGNMENTS

RESULT 1  
US-08-481-658B-65/c  
; Sequence 65, Application US/08481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,658B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/COCKET NUMBER: D-0021.3E  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-481-658B-65

Query Match 0.3%; Score 82.6; DB 4; Length 105;  
Best Local Similarity 86.7%; Pred. No. 2.5e-08;

```
Matches 91; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 18973 ATCCGAGCAGCTTTGGAGCGCAAGCGGACAGATCAGAGGTCTAGGAGTTTGAGACGAGC 19032
Db 105 ATCCGAGCAGCTTTGGAGCGCGAGCGTGTGTGATCACAAGGTCTAGGAGTTTGAGACGAGC 46
QY 19033 CTGACCAACATGGTGAACACCTGTCTCTACTAACAATAACAAAA 19077
Db 45 CTGGCCATATGGTGAACACCTGTCTCTACTAACAAGATGTAATAAAA 1

RESULT 2
US-08-477-504A-65/c
; Sequence 65, Application US/08477504A
; Patent No. 5972353
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,504A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-477-504A-65

Query Match 0.3%; Score 82.6; DB 4; Length 105;
Best Local Similarity 86.7%; Pred. No. 2.5e-08;
Matches 91; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 18973 ATCCGAGCAGCTTTGGAGCGCAAGCGGACAGATCAGAGGTCTAGGAGTTTGAGACGAGC 19032
Db 105 ATCCGAGCAGCTTTGGAGCGCGAGCGTGTGTGATCACAAGGTCTAGGAGTTTGAGACGAGC 46
QY 19033 CTGACCAACATGGTGAACACCTGTCTCTACTAACAATAACAAAA 19077
Db 45 CTGGCCATATGGTGAACACCTGTCTCTACTAACAAGATGTAATAAAA 1

RESULT 3
```

```
US-08-486-756A-65/c
; Sequence 65, Application US/08486756A
; Patent No. 5981711
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,756A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,190
; FILING DATE: 15-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-486-756A-65

Query Match 0.3%; Score 82.6; DB 4; Length 105;
Best Local Similarity 86.7%; Pred. No. 2.5e-08;
Matches 91; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 18973 ATCCGAGCAGCTTTGGAGCGCAAGCGGACAGATCAGAGGTCTAGGAGTTTGAGACGAGC 19032
Db 105 ATCCGAGCAGCTTTGGAGCGCGAGCGTGTGTGATCACAAGGTCTAGGAGTTTGAGACGAGC 46
QY 19033 CTGACCAACATGGTGAACACCTGTCTCTACTAACAATAACAAAA 19077
Db 45 CTGGCCATATGGTGAACACCTGTCTCTACTAACAAGATGTAATAAAA 1

RESULT 4
US-08-485-862B-65/c
; Sequence 65, Application US/08485862B
; Patent No. 5989838
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
```

CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485.862B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-485-862B-65

Query Match 0.3%; Score 82.6; DB 4; Length 105;  
Best Local Similarity 86.7%; Pred No. 2.5e-08;  
Matches 91; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
QY 18973 ATCCGAGCATTGGGAGGCCGACAGATCAGAGGTCAGAGTTTGAGACCAGC 19032  
DB 105 ATCCGAGCATTGGGAGGCCGAGCTGTGATCACAAGGTCAGAGTTTGAGACCAGC 46  
QY 19033 CTGACCAACATGGTGAACCCCTGTCTCTACTACAAAATACAAA 19077  
DB 45 CTGGCCAATATGGTGAACCCCTGTCTCTACTAAAGATGTAAAAA 1

Query Match 0.3%; Score 82.6; DB 4; Length 105;  
Best Local Similarity 86.7%; Pred No. 2.5e-08;  
Matches 91; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
QY 18973 ATCCGAGCATTGGGAGGCCGACAGATCAGAGGTCAGAGTTTGAGACCAGC 19032  
DB 105 ATCCGAGCATTGGGAGGCCGAGCTGTGATCACAAGGTCAGAGTTTGAGACCAGC 46  
QY 19033 CTGACCAACATGGTGAACCCCTGTCTCTACTACAAAATACAAA 19077  
DB 45 CTGGCCAATATGGTGAACCCCTGTCTCTACTAAAGATGTAAAAA 1

RESULT 5  
US-08-787-739-65/C  
; Sequence 65, Application US/08787739  
; Patent No. 6027887  
; GENERAL INFORMATION:  
; APPLICANT: zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 96  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 369 Pine Street, Suite 610  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/787,739  
FILING DATE: 24-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,049  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/486,756  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/481,658  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,862  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,863  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/487,077  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-981-2034  
TELEFAX: 415-981-0332  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-787-739-65

Query Match 0.3%; Score 82.6; DB 5; Length 105;  
Best Local Similarity 86.7%; Pred. No. 2.5e-08;  
Matches 91; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
QY 18973 ATCCGAGCATTGGGAGGCCGACAGATCAGAGGTCAGAGTTTGAGACCAGC 19032  
DB 105 ATCCGAGCATTGGGAGGCCGAGCTGTGATCACAAGGTCAGAGTTTGAGACCAGC 46  
QY 19033 CTGACCAACATGGTGAACCCCTGTCTCTACTACAAAATACAAA 19077  
DB 45 CTGGCCAATATGGTGAACCCCTGTCTCTACTAAAGATGTAAAAA 1

RESULT 6  
US-08-454-557C-91/C  
; Sequence 91, Application US/08454557C  
; Patent No. 5830670  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; APPLICANT: Wands, Jack R.  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/454,557C  
;; FILING DATE: 30-MAY-1995  
;; CLASSIFICATION: 514  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Ludwig, Steven R.  
;; REGISTRATION NUMBER: 36,203  
;; REFERENCE/DOCKET NUMBER: 0609.3840003  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 371-2600  
;; TELEFAX: (202) 371-2540  
;; INFORMATION FOR SEQ ID NO: 91:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 84 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: both  
;; TOPOLOGY: both  
US-08-454-557C-91

Query Match 0.2%; Score 68.6; DB 3; Length 84;  
Best Local Similarity 89.2%; Pred. No. 1.5e-05;  
Matches 74; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
  
Qy 18963 CACGCTGTATCCAGCACACCTTTGGGAGGCCAAGCGGACAGATCAGAGTCTAGGAGTT 19022  
Db 83 CACGCTGTATCCAGCACACCTTTGGGAGGCCAAGCGGACAGATCAGAGTCTAGGAGTT 24  
  
Qy 19023 TGAGACACGCTGACCAACATGG 19045  
Db 23 CGACACGCTGATGACATGG 1

RESULT 7  
US-08-426D-91/c  
; Sequence 91, Application US/08340426D  
; Patent No. 5948634  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; APPLICANT: Wands, Jack R.  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; TITLE OF INVENTION: of Alzheimer's Disease  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/340,426D  
; FILING DATE: 14-NOV-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, Steven R.  
; REGISTRATION NUMBER: 36,203  
; REFERENCE/DOCKET NUMBER: 0609.3840002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 91:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 84 base pairs

;; TYPE: nucleic acid  
;; STRANDEDNESS: both  
;; TOPOLOGY: both  
US-08-340-426D-91  
  
Query Match 0.2%; Score 68.6; DB 4; Length 84;  
Best Local Similarity 89.2%; Pred. No. 1.5e-05;  
Matches 74; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
  
Qy 18963 CACGCTGTATCCAGCACACCTTTGGGAGGCCAAGCGGACAGATCAGAGTCTAGGAGTT 19022  
Db 83 CACGCTGTATCCAGCACACCTTTGGGAGGCCAAGCGGACAGATCAGAGTCTAGGAGTT 24  
  
Qy 19023 TGAGACACGCTGACCAACATGG 19045  
Db 23 CGACACGCTGATGACATGG 1

RESULT 8  
US-08-450-673C-91/c  
; Sequence 91, Application US/08450673C  
; Patent No. 5948888  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; APPLICANT: Wands, Jack R.  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; TITLE OF INVENTION: of Alzheimer's Disease  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/450,673C  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, Steven R.  
; REGISTRATION NUMBER: 36,203  
; REFERENCE/DOCKET NUMBER: 0609.3840004  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 91:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 84 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
US-08-450-673C-91

Query Match 0.2%; Score 68.6; DB 4; Length 84;  
Best Local Similarity 89.2%; Pred. No. 1.5e-05;  
Matches 74; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
  
Qy 18963 CACGCTGTATCCAGCACACCTTTGGGAGGCCAAGCGGACAGATCAGAGTCTAGGAGTT 19022  
Db 83 CACGCTGTATCCAGCACACCTTTGGGAGGCCAAGCGGACAGATCAGAGTCTAGGAGTT 24  
  
Qy 19023 TGAGACACGCTGACCAACATGG 19045  
Db 23 CGACACGCTGATGACATGG 1



RESULT 9  
PCT-US95-17111A-91/c  
; Sequence 91, Application PC/TUS9517111A  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; APPLICANT: Wands, Jack R.  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and  
; TITLE OF INVENTION: Detection of Alzheimer's Disease  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/17111A  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/340,426  
; FILING DATE: 14-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, Steven R.  
; REGISTRATION NUMBER: 36,203  
; REFERENCE/DOCKET NUMBER: 0609.3840002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 91:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 84 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
PCT-US95-17111A-91  
  
Query Match 0.2%; Score 68.6; DB 6; Length 84;  
Best Local Similarity 89.2%; Pred. No. 1.5e-05;  
Matches 74; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
  
Qy 18963 CACGCTGTATCCAGCACCTTTGGAGGCCAAGCGGACAGATCACGAGGTCAGAGTT 19022  
Db 83 CACGCTGTATCCAGCACCTTTGGAGGCTGAGCGGCGGATCACGAGGTCAGAGTT 24  
  
Qy 19023 TGAGACCAAGCTGACCAACATGG 19045  
Db 23 CGACACCAAGCTGATGAACATG 1  
  
RESULT 10  
US-08-481-658B-65  
; Sequence 65, Application US/08481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California

; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,658B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.3E  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-481-658B-65  
  
Query Match 0.2%; Score 65.6; DB 4; Length 105;  
Best Local Similarity 84.3%; Pred. No. 6.9e-05;  
Matches 86; Conservative 0; Mismatches 14; Indels 2; Gaps 1;  
  
Qy 27880 TTTTATATTTTGTAGAGACGGGTTTCAGCATGTTGGTCAAGCTTGTCTCAAACTCC 27939  
Db 2 TTTTATATCTTTAGTAGACAGAGGTTTCACCATATTTGGCCAGCGCTGCTCTCAAACTCC 61  
  
Qy 27940 TGACCTCAGGTATCCACCACCTCGGCTCCGAAAGTTG 27981  
Db 62 TGACCT--TGTATCCACGAGCTCGGCTCCCAAGTGCTG 101  
  
RESULT 11  
US-08-477-504A-65  
; Sequence 65, Application US/08477504A  
; Patent No. 5972353  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/477,504A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:

Query Match 0.28; Score 65.6; DB 4; Length 105;



Db 80 AACACACACACA 91

Search completed: June 20, 2000, 18:07:04  
Job time: 537154 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2000, 17:56:59 ; Search time 29135.9 Seconds  
(without alignments)  
-601.018 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_196000\_214000  
Perfect score: 18001  
Sequence: 1 CTAAGAGTATTTTAAAC.....TGGACATGCTGTCTTCCTC 18001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 370290

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_on.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pl1.\*
- 8: gb\_pl2.\*
- 9: gb\_pr1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_v1.\*
- 17: em\_fun.\*
- 18: em\_hum1.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_on.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pl.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_v1.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_ba1.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*
- 44: gb\_htg6.\*

- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*
- 50: gb\_pl3.\*
- 51: gb\_pr5.\*
- 52: gb\_htg8.\*
- 53: gb\_htg9.\*
- 54: gb\_htg10.\*
- 55: gb\_htg11.\*
- 56: gb\_htg12.\*
- 57: gb\_htg13.\*
- 58: gb\_htg14.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
c 1	88.8	0.5	108	10 HSLDLRN2	X05250 Human LDL-r
c 2	87.4	0.5	108	11 HSU67803	U67803 Human small
c 3	84.4	0.5	103	9 HUMALCE221	M87896 Human carc
5	84	0.5	108	10 HSLDLRN2	X05250 Human LDL-r
c 6	82.2	0.5	107	10 HSLDLI12	X05248 Human LDL-r
c 7	79.8	0.4	108	10 HUMALCE162	M87924 Human carc
c 8	79.8	0.4	108	10 HSLDLRD1	X05249 Human LDL-r
c 9	79.8	0.4	108	10 HSLDLRD2	X05251 Human LDL-r
c 10	79.8	0.4	108	10 HSLDLRD2	X05251 Human LDL-r
c 11	79.2	0.4	103	13 HS8IC8R	X57789 Human sequ
c 12	77.8	0.4	108	11 HSU67804	U67804 Human small
c 13	77.2	0.4	91	13 HUMUT8164A	L30244 Human STS U
c 14	77.2	0.4	110	11 HSU67807	U67807 Human small
c 15	76.2	0.4	108	11 HSU67808	U67808 Human small
c 16	75.6	0.4	103	13 HS8IC8R	X57789 Human sequ
c 17	74.4	0.4	110	9 HUMALCE43	M87900 Human carc
c 18	74.2	0.4	107	9 HUMALCE162	M87924 Human carc
c 19	74.2	0.4	108	9 HUMD1D03M5	D16965 Human HepG2
c 20	74	0.4	103	9 HUMALCE221	M87896 Human carc
c 21	73.4	0.4	110	11 HSU67807	U67807 Human small
c 22	73.2	0.4	104	9 HUMALCE272	M87899 Human carc
c 23	72.4	0.4	108	10 HSLDLI12	X05248 Human LDL-r
c 24	72	0.4	90	9 HUMDLRFL	K03555 Human low d
c 25	72	0.4	106	13 G32743	G32743 A009P31 Hum
c 26	71.2	0.4	108	13 G43535	G43535 WIAF-2393-S
c 27	70.4	0.4	97	9 HUMDLRA2	M14180 Human low d
c 28	70.4	0.4	106	13 G32743	G32743 A009P31 Hum
c 29	69.8	0.4	90	9 HUMDLRFL	K03555 Human low d
c 30	69.8	0.4	104	9 HUMALCE272	M87899 Human carc
c 31	69.4	0.4	101	10 S79560	S79560 HRX (intron
c 32	69.2	0.4	97	9 HUMDLRA1	M14178 Human low d
c 33	69	0.4	108	11 HSU67803	U67803 Human small
c 34	68.8	0.4	107	13 HUMUT7692A	L30306 Human STS U
c 35	68.2	0.4	109	11 HSU67806	U67806 Human small
c 36	68	0.4	108	13 G43535	G43535 WIAF-2393-S
c 37	67.8	0.4	108	3 AF185109S1	AF185109 Lasiorhin
c 38	67.8	0.4	108	9 HUMD1D03M5	D16965 Human HepG2
c 39	67.6	0.4	100	13 HUMUT931A	L31299 Human STS U
c 40	67.2	0.4	80	9 HUMBRKFAE	M36135 Human alpha
c 41	67.2	0.4	97	9 HUMDLRA2	M14180 Human low d
c 42	67	0.4	91	13 HUMUT8164A	L30244 Human STS U
c 43	67	0.4	94	9 HUMDLRA1	M14178 Human low d
c 44	66.8	0.4	94	9 HUMHGAL	M13479 Human alpha
c 45	66.8	0.4	100	11 HSU67848	U67848 Human beta-

ALIGNMENTS



```

REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT See X05252 for deletion junction
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES
source Location/Qualifiers
1..108
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..108
/note="intron XIV fragment"
BASE COUNT 28 a 23 c 39 g 18 t
ORIGIN
Query Match 0.5%; Score 84; DB 10; Length 108;
Best Local Similarity 86.1%; Pred. No. 8.4e-05;
Matches 93; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Qy 1765 AAAAAAATAGCTGGCGAGGTGCCAGGACCTGTATCCAGCTACTGGGAGGCTGAG 1824
Db 1 AAAAAATAGCCAGCGTGTGGCAGGTGCTGTATCCAGCTACTGGGAGGCTGAG 60
Qy 1825 GGAGCAATACATGTAACTCAGGAGGAGGAGTGTAGTGTAGTGAGTGAG 1872
Db 61 GCAGGAGAATTGCTTGAACCCAGGAGGAGGAGTGTGAGTGAGCGGAG 108
RESULT 5
HSLDL12 108 bp DNA PRI 20-MAY-1992
LOCUS Human LDL-receptor gene Intron 12 fragment (normal gene) LDL - low
DEFINITION density lipoprotein.
ACCESSION X05248
VERSION X05248.1 GI:34334
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor;
repetitive sequence.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT See X05249 for deletion junction
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES
source Location/Qualifiers
1..108
/organism="Homo sapiens"
/db_xref="taxon:9606"
complement(<1..65)
/note="Alu repeat"
1..108
/note="intron XII fragment"
BASE COUNT 21 a 38 c 20 g 29 t
ORIGIN
Query Match 0.5%; Score 83; DB 10; Length 108;
Best Local Similarity 86.0%; Pred. No. 0.00012;
Matches 92; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Qy 13476 TTGGTCACTCAACCTTGGCTCTTGGGTTCAAGCAATTATCTCGCTCAGCCTCCCTTA 13535

```

```

Db 2 TCGCCTCACCACAACCTCTGCTCTCTGGGTTCAAACCAATTTTCTGCTCAGCCTCCCTTA 61
Qy 13536 GTACAGGGTCTACAGGATGTGCCACACACACCGCGCTGTTTTGTA 13582
Db 62 GTAGCTGGGATTACAAGCATGTGCCACACACCGCGCTGATTTTGTGA 108
RESULT 6
HUMALCE162/c 107 bp ss-RNA PRI 15-APR-1994
LOCUS Human carcinoma cell-derived Alu RNA transcript, clone CE162.
ACCESSION M87924
VERSION M87924.1 GI:174871
KEYWORDS Alu repeat.
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 107)
AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
post-transcriptional selection of master sequences
JOURNAL J. Mol. Biol. (1992) in press
FEATURES
Location/Qualifiers
source
1..107
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="NterazDI"
/dev_stage="embryo"
/sex="male"
/tissue_type="carcinoma"
BASE COUNT 28 a 30 c 35 g 14 t
ORIGIN
Query Match 0.5%; Score 82.2; DB 9; Length 107;
Best Local Similarity 87.4%; Pred. No. 0.00017;
Matches 90; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Qy 14973 TTTTCTTTGACTGAGTCTGCTGCTGCTCAGAGGTGAGTGCAGTGTGGATCTCGC 15032
Db 107 TTTTCTGAGACGGAGTCTGCTGCTGCTCGCCAGGTGAGTGCAGTGTGGATCTCGC 48
Qy 15033 TCACTGCAACCTCTGCCTCCGCGGTTCAGTGATTTCTCTGCC 15075
Db 47 TCACTGCAAGCTCCGCCCTCCCGGGTTACAGCCATTCCTCTGCC 5
RESULT 7
HSLDLRDI 108 bp DNA PRI 20-MAY-1992
LOCUS Human LDL-receptor mutated gene with intron 12 deletion junction.
DEFINITION X05249
ACCESSION X05249.1 GI:34335
VERSION X05249.1 GI:34335
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT *source: hypercholesterol aemia
See X05248 for corresponding normal gene sequence
In the defective LDL-receptor gene the deletion occurred between two
alu-repetitive sequences, that are in the same direction, the
deletion eliminates exons 13 and 14 and changes the reading frame

```

```
of the resulting spliced mRNA.
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
Location/Qualifiers
1. .108
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="blood leukocytes from a patient with familial"
misc_feature
1. .108
/notes="deletion junction region intron 12/ intron 15"
BASE COUNT 20 a 40 c 20 g 28 t
ORIGIN
Query Match 0.4%; Score 79.8; DB 10; Length 108;
Best Local Similarity 84.1%; Pred. No. 0.00042;
Matches 90; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 15730 TTGGCTGCTACAGCATGCTGCTGGTTCAGTGATTCCTGACTCAGCTCCCGA 15789
Db 2 TCGCCTCACCACACACCTGCTGCTGCTGGTTCAAACCATTTTCTGCTCAGCTCCCGA 61
QY 15790 GTAGCTGGGATTACAGGCATGCTACCATGCTGCTGGTAAATTTTGT 15836
Db 62 GTAGCTGGGATTACAGGCATGCTGCTGGTAAATTTTGT 108
RESULT 8
HSLDLR1/c 108 bp DNA PRI 20-MAY-1992
LOCUS Human LDL-receptor mutated gene with intron 12 deletion junction.
DEFINITION X05249
VERSION X05249.1 GI:34335
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
Williamson,R. and Humphries,S.
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT *source: hypercholesterol aemia
See X05248 for corresponding normal gene sequence
In the defective LDL-receptor gene the deletion occurred between two
alu-repetitive sequences, that are in the same direction, the
deletion eliminates exons 13 and 14 and changes the reading frame
of the resulting spliced mRNA.
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
Location/Qualifiers
1. .108
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="blood leukocytes from a patient with familial"
intron
1. .108
/notes="intron XIV fragment"
BASE COUNT 28 a 20 c 40 g 20 t
ORIGIN
Query Match 0.4%; Score 79.8; DB 10; Length 108;
Best Local Similarity 84.1%; Pred. No. 0.00042;
Matches 90; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 1765 AAAAAAATTAGCTGGCGAGGTGCCAGGACCTGCTAATCCAGCTACTCGGAGGCTGAG 1824
Db 1 AAAAAAATTAGCAGCGCTGGTGGCAGGTGCTGTAATCCAGCTACTCGGAGGCTGAG 60
QY 1825 GGAGGACAATCATTGAACCTCAGGAGCGAGAGGTTGTAGTGAGCTGA 1871
Db 61 GCAGGAAAATGGTTTGAACCCAGGAGCGAGAGGTTGTGTGAGGCGA 107
RESULT 10
HSLDLR2/c 108 bp DNA PRI 20-MAY-1992
LOCUS Human LDL-receptor mutated gene with intron 14 deletion junction.
DEFINITION X05251
VERSION X05251.1 GI:34336
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
Williamson,R. and Humphries,S.
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT *source: hypercholesterol aemia
```



See X05250 for corresponding normal gene sequence  
 In the defective LDL-receptor gene the deletion occurred between two  
 alu-repetitive sequences, that are in the same direction, the  
 deletion eliminates exons 13 and 14 and changes the reading frame  
 of the resulting spliced mRNA.  
 Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

```

FEATURES
  source
    1. .108
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /cell_type="blood leukocytes from a patient with familial"
  intron
    1. .108
    /note="intron XIV fragment"
  BASE COUNT      28 a   20 c   40 g   20 t
  ORIGIN

Query Match      0.4%; Score 79.8; DB 10; Length 108;
Best Local Similarity 84.1%; Pred. No. 0.00042;
Matches 90; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 15730 TTGGCTCATTCACACCTCGCTCCTGGGTCAAGTGATTCCTGACTCAGCCTCCCGA 15789
Db 107 TCGCCTCACCACACCTCGCTCCTGGGTCAAAACCATTTCTGCTCAGCCTCCCGA 48
Qy 15790 GTAGCTGGGATTACAGGACGATCACCATGCCCTGGGTAAATTTTGT 15836
Db 47 GTAGCTGGGATTACAGGACGCTGCCACGCCCGCTAAATTTTGT 1

RESULT 11
HS81C8R/c      103 bp      DNA      STS      05-SEP-1991
LOCUS          Human sequence tagged site 81C8R DNA from 19q13.
DEFINITION     X57789
ACCESSION      X57789.1 GI:23938
VERSION         STS; myotonic dystrophy.
KEYWORDS        human.
SOURCE          Homo sapiens
ORGANISM        Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
                  Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 103)
AUTHORS        Aldridge,F.L.
TITLE          Direct Submission
JOURNAL        Submitted (12-FEB-1991) F.L. Aldridge, ICI Pharmaceuticals,
                  Alderley Park, Macclesfield, Cheshire, SK10 4TG, UK
REFERENCE      2 (bases 1 to 103)
AUTHORS        Butler,R., Riley,J.H., Ogilvie,D.J., Anand,R., Buxton,J.,
                  Davies,J., Johnson,K. and Markham,A.F.
TITLE          Two sequence-tagged sites defining the ends of a 380 kb YAC clone
                  from 19q13
JOURNAL        Nucleic Acids Res. 19 (17), 4787 (1991)
MEDLINE        91367697
COMMENT        See also X57788 for STS 81C8L.
FEATURES
  source
    1. .103
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="19q13"
    /germline
    /clone_lib="YAC library: ICI"
    /clone="81C8"
  BASE COUNT      29 a   28 c   23 g   22 t   1 others
  ORIGIN

Query Match      0.4%; Score 79.2; DB 13; Length 103;
Best Local Similarity 86.1%; Pred. No. 0.00054;
Matches 87; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 1656 TACCTGTAATCCACGACTTTGGGAGACTGAGGTGGGTGGATCACTTGAGTCAGGAGTT 1715
Db 102 TGCCTATAATTCGCACTTTGGGAGGTCGAGGTGGGTGGATCACTTAAGGTCAGGAGTT 43

```

```

Qy 1716 CAAACACGAGCTGGCCACATGGTGAACACCATCTCTACT 1756
Db 42 CTTGACCAGCTGGCCACATGGTGAACACCATCTCTACT 2

RESULT 12
HSU67804/c      108 bp      RNA      PRI      01-AUG-1997
LOCUS          Human small cytoplasmic Alu transcript.
DEFINITION     U67804
ACCESSION      U67804.1 GI:2289918
VERSION         Alu.
KEYWORDS        human.
SOURCE          Homo sapiens
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                  Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 108)
AUTHORS        Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE          cDNAs derived from primary and small cytoplasmic Alu (scAlu)
                  transcripts
JOURNAL        J. Mol. Biol. 271 (2), 222-234 (1997)
MEDLINE        97415756
REFERENCE      2 (bases 1 to 108)
AUTHORS        Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE          Direct Submission
JOURNAL        Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The
                  Children's Hospital of Philadelphia, 1004F Abramson Research
                  Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA
FEATURES
  source
    1. .108
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="TscAlu3"
    /rpt_family="Alu"
    /rpt_type="dispersed"
  repeat_region
    26 a   38 c   26 g   18 t
  BASE COUNT      26 a   38 c   26 g   18 t
  ORIGIN

Query Match      0.4%; Score 77.8; DB 11; Length 108;
Best Local Similarity 87.6%; Pred. No. 0.00091;
Matches 85; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 15143 GTAGAGATGGAGTTTCGCCGTGTAGCCAGGATGTCGATCTCCTGACCTCGTGATCC 15202
Db 97 GGAAGACGGGGTTTCACCATGTTAGCCAGGATGTCGATCTCCTGACCTTGATCC 38
Qy 15203 ACCGGCTCGGCCCTCCAAAGTGGGATTACAGGC 15239
Db 37 TCCCGCTTTGGCCTTCCAAAGTGGGATTACAGGC 1

RESULT 13
HUMUT8164A/c      91 bp      DNA      STS      29-DEC-1994
LOCUS          Human STS UT8164, 5' primer bind, sequence tagged site.
DEFINITION     L30244
ACCESSION      L30244.1 GI:605447
VERSION         STS; PCR primer; STS sequence; microsatellite DNA; microsatellite
KEYWORDS        marker; sequence tagged site; tetranucleotide repeat.
SOURCE          Homo sapiens DNA.
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                  Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 91)
AUTHORS        Gerken,S.C., Matsunami,N., Plaetke,R., Albertsen,H., Ballard,L.,
                  Mellis,F., Lawrence,E., Moore,M., Holik,P.R., Carlson,M., Zhao,X.,
                  Robertson,M., Bradley,P., Elsner,T., Tingey,A., Lalouel,J.-M. and
                  White,F.
TITLE          Genetic and physical mapping of simple sequence repeat containing

```









CC markers which have been isolated using the primers represented in  
CC X09121-X10268. The base occupying the polymorphic site is indicated by  
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC spherocytosis, von Willebrand's disease, tuberculous sclerosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.

Query Match 0.4%; Score 71.8; DB 1; Length 108;  
Best Local Similarity 84.9%; Pred. No. 0.019;  
Matches 79; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

Qy 15834 TGATATTTAGTAGAGATGGGTTTACCATGTCACAGAGCTGGTCTCAAACTCTGAC 15893  
Dd 1 TTTCTCTTTTGTAGATAGATTTTTTCTCTGTGGCAGGATGCTCTCGAACTCTGAC 60

Qy 15894 CTCAGTGCATCCGCTGAGCTCCCAAAA 15926  
Dd 61 TTCAGTGCATCCGCTGAGCTCCCAAAA 93

RESULT 3  
X12095/C  
ID X12095 standard; DNA; 108 BP.  
AC X12095;  
DT 30-MAR-1999 (first entry)  
DE Human biallelic polymorphic DNA fragment TIGR-A003M18a.  
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;  
KW detection; phenotypic typing; characteristic; infection; hereditary;  
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;  
KW treatment; marker; ss.  
OS Homo sapiens.  
PN W09820165-A2.  
PD 14-MAY-1998.  
PF 05-NOV-1997; U20313.  
PI (WHED) WHITEHEAD INST BIOMEDICAL RES.  
PI Hudson T, Lander ES, Wang D;  
DR WPI; 98-286974/25.  
PT New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease

CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
CC markers which have been isolated using the primers represented in  
CC X09121-X10268. The base occupying the polymorphic site is indicated by  
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC spherocytosis, von Willebrand's disease, tuberculous sclerosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.

Query Match 0.4%; Score 68.6; DB 1; Length 108;  
Best Local Similarity 83.8%; Pred. No. 0.055;  
Matches 88; Conservative 1; Mismatches 15; Indels 1; Gaps 1;

Qy 6421 TAATCCCGACAC-TTTGGAGGCGGAGGAGGATCATGAGTCAGAAATTCAGA 6479  
Dd 105 TAATCCCGACAC-TTTGGAGGCGGAGGAGGATCATGAGTCAGAAATTCAGA 46

Qy 6480 CCAGCTGGCCAAATGGTGAACCTCATCTTACTATAAATACA 6524  
Dd 45 CCATCTCTGGCCCAACAYAGGAAACCTCATCTTACTATAAAGACA 1

RESULT 4  
T24892/C  
ID T24892 standard; cDNA to mRNA; 100 BP.  
AC T24892;  
DT 05-NOV-1996 (first entry)  
DE Human gene signature HUMGS06998.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN W09514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATSU) MATSUBARA K.  
PA (OKUBU) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI; 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues

PS Claim 1; Page 1720; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.

SQ Sequence 100 BP; 28 A; 23 C; 28 G; 37 T;

CC markers which have been isolated using the primers represented in  
CC X09121-X10268. The base occupying the polymorphic site is indicated by  
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC spherocytosis, von Willebrand's disease, tuberculous sclerosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.

Query Match 0.4%; Score 68.6; DB 1; Length 108;  
Best Local Similarity 83.8%; Pred. No. 0.055;  
Matches 88; Conservative 1; Mismatches 15; Indels 1; Gaps 1;

Qy 6421 TAATCCCGACAC-TTTGGAGGCGGAGGAGGATCATGAGTCAGAAATTCAGA 6479  
Dd 105 TAATCCCGACAC-TTTGGAGGCGGAGGAGGATCATGAGTCAGAAATTCAGA 46

Qy 6480 CCAGCTGGCCAAATGGTGAACCTCATCTTACTATAAATACA 6524  
Dd 45 CCATCTCTGGCCCAACAYAGGAAACCTCATCTTACTATAAAGACA 1

RESULT 4  
T24892/C  
ID T24892 standard; cDNA to mRNA; 100 BP.  
AC T24892;  
DT 05-NOV-1996 (first entry)  
DE Human gene signature HUMGS06998.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN W09514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATSU) MATSUBARA K.  
PA (OKUBU) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI; 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues

PS Claim 1; Page 1720; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.

SQ Sequence 100 BP; 28 A; 23 C; 28 G; 37 T;

Query Match	0.4%;	Score 65.6;	DB 1;	Length 100;	
Best Local Similarity	77.8%;	Pred. No. 0.14;			
Matches	77;	Conservative	0;	Mismatches	22;
				Indels	0;
				Gaps	0;
QY	14968	TTTTTTTTTTTTTTGACGTAGCTCTGCTCTCACAGGCTGGAGTCGAGTCGGTCGATC	15027		
DB	100	TTTGTGTTGTTTCAAACAGAGTGTCACTCTGTCTACCCAGCGNGAGTGC AANGTGCAATC	41		
QY	15028	TCGGCTCACGTCAACCTCTGCCTCCCGGGTTCAAGTGAT	15066		
DB	40	TCAGCTNATTTGCAAAATTTCTGCCTCCCAAGGTTCAAGCGAT	2		
RESULT	5				
T26213/C					
ID	T26213	standard;	cdna to mRNA;	103 BP.	
AC	T26213;				
AT	13-NOV-1996	(first entry)			
DE	Human gene signature HUMGS08452.				
KW	Gene signature; messenger RNA; mRNA;	relative abundance;	frequency;		
DE	human; cloning; mapping; non-biased library;	diagnosis;	detection;		
KW	cell typing; abnormal cell function;	ss.			
OS	Homo sapiens.				
PN	W09514/772-A1.				
PD	01-JUN-1995.				
PF	11-NOV-1994;	J01916.			
PR	12-NOV-1993;	JP-355504.			
PA	(MATS/) MATSUBARA K.				
PA	(OKUB/) OKUBO K.				
PI	Matsubara K, Okubo K;				
PI	WPI; 95-206931/27.				
PT	Identifying gene signatures in 3'-directed human cDNA library - e.g.				
PT	for diagnosis of abnormal cell function, by preparing cDNA that				
PT	reflects relative abundance of corresp. mRNA in specific human				
PT	tissues				
PS	Claim 1; Page 2029; 2245pp; Japanese.				
CC	A single-stranded DNA (or its complementary strand or the corresp.				
CC	double-stranded DNA) which comprises one of the 7837 "GS" sequences				
CC	given in T19001-T26837 and which is able to hybridise to part of				
CC	human genomic DNA, cDNA or mRNA is claimed. The GS (gene Signature)				
CC	sequences were obtained from 3'-directed cDNA libraries prepared				
CC	from various human tissues; synthesis of cDNA was initiated from the				
CC	3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-				
CC	untranslated sequence is unique to a particular mRNA species, almost				
CC	all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library				
CC	is constructed so as to reflect accurately the relative abundance of				
CC	different mRNAs in the particular tissue from which it was derived.				
CC	The appearance frequency of a given GS in a cDNA library can be				
CC	determined (esp. using primers and probes derived from the GS				
CC	sequences) as a means of diagnosing abnormal cell function or for				
CC	recognising different cell types.				
SQ	Sequence	103 BP;	33 A;	25 C;	23 T;
Query Match	0.4%;	Score 65.2;	DB 1;	Length 103;	
Best Local Similarity	77.5%;	Pred. No. 0.16;			
Matches	79;	Conservative	0;	Mismatches	23;
				Indels	0;
				Gaps	0;
QY	8118	TTTCTTTTCTTTCTGTACAGGGTCTTCTCTATTGTCCTAGGCTGAGTCAGTGGTGCA	8177		
DB	102	TTTTTTTTTTCTTAAAGACATGTTCTTACTCTGTGCCAGGCTGGAGTCAGTGGTGCA	43		
QY	8178	TCATAGCTCACTCAGCGCTTGAACCTCAGGCTCAAGCAATC	8219		
DB	42	TCATAGCTCACTGTACACCAACCAACTCTGGACTCAAGTGATC	1		
RESULT	6				
T66081					
ID	T66081	standard;	DNA;	92 BP.	
AC	T66081;				

PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues

PS Claim 1: Page 1720: 2245pp: Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 100 BP; 28 A; 22 C; 25 G; 22 T;

Query Match 0.4%; Score 64; DB 1; Length 100;  
Best Local Similarity 76.8%; Pred. No. 0.24;  
Matches 76; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1833 ATCACTTGAACTCAGGAGGAGGTGTAGTGAGCTGAGATCGACCACTGCACCTCCAG 1892

Db 2 ATCCCTTGAACTCGGAGGAGAAATTCGATNAGCTGAGATTGCACCTGCACTCCNG 61

QY 1893 CCTGGTGACAGCAAGACTCCATTTAAATAATAATA 1931

Db 62 CCTGGTGACAGAGTGACACTCTCTTTGAAACAAACA 100

RESULT 8

T25009/c  
ID T25009 standard; cDNA to mRNA; 108 BP.  
AC T25009;  
DT 07-NOV-1996 (first entry)  
DE Human gene signature HUMGS07131.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN WO9514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
PI Matsubara K. Okubo K;  
DR WPI; 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues

PS Claim 1: Page 1748: 2245pp: Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.

SQ Sequence 108 BP; 34 A; 31 C; 26 G; 15 T;

Query Match 0.4%; Score 64; DB 1; Length 108;  
Best Local Similarity 74.5%; Pred. No. 0.24;  
Matches 79; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 17285 TTGTGTTGTTGTTTGTGATAGAGTCTTCCTCTCATTCAGGTGAGTGCAGTGG 17344

Db 107 TTGNTGTGTTGTTGTTTCAACAGGGCTTGCTCTCTCACTCAGGTGGATNACAGTGG 48

QY 17345 CATGATCTCAGCTCAGTCAGCCCTCCGCCGTTCAAGAGAT 17390

Db 47 CGTGACCATGGCTCACTGCAGCCCTTGCCCTCATGGGCTCAGGCGAT 2

RESULT 9

T26828  
ID T26828 standard; cDNA to mRNA; 108 BP.

AC T26828;

DT 14-NOV-1996 (first entry)

DE Human gene signature HUMGS09078.

KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;

KW human; cloning; mapping; non-biased library; diagnosis; detection;

KW cell typing; abnormal cell function; ss.

OS Homo sapiens.

PN WO9514772-A1.

PD 01-JUN-1995.

PF 11-NOV-1994; J01916.

PR 12-NOV-1993; JP-355504.

PA (MATS/) MATSUBARA K.

PA (OKUB/) OKUBO K.

PI Matsubara K. Okubo K;

DR WPI; 95-206931/27.

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.

PT for diagnosis of abnormal cell function, by preparing cDNA that

PT reflects relative abundance of corresp. mRNA in specific human

PT tissues

PS Claim 1: Page 2182: 2245pp: Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in T19001-T26837 and which is able to hybridise to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost

CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library

CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be

CC determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.

SQ Sequence 108 BP; 18 A; 33 C; 23 G; 28 T;

Query Match 0.4%; Score 63.8; DB 1; Length 108;

Best Local Similarity 88.3%; Pred. No. 0.26;

Matches 68; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 15182 GATCTCTGTGACCTCGTGATCCACCGCTCGGCTCCCAAGTCTGGATTCAGGCAT 15241

Db 1 GATCTCTGTGACCTCGTGATCCCGCCGCTGCTCCCATAGTCTGGGNTTACAGGCAT 60

QY 15242 GGGCCACACCGCTGGC 15258

Db 61 GAGCCACACCGCCGGC 77

RESULT 10

V11611/c

ID V11611 standard; cDNA; 106 BP.







CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 87 BP; 35 A; 21 C; 16 G; 13 T

Query Match 0.3%; Score 58.4; DB 1; Length 87;  
Best Local Similarity 79.1%; Pred. No. 1.5;  
Matches 68; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 15819 TGCCTGGGTAATTTTGTATTTTAGTAGAGATGGGGTTTCACCATGTTGACCAGGCTGG 15878

Db 86 TTCTGGCTNATTTCTGTAATTTTTGTAAGATGGGGTTTCGCCATGTTCTTCGGGCTGG 27

QY 15879 TCTCAAACTCCTGACCTCAAGTGATC 15904

Db 26 TTTTAAACTCCTGGGNTCAAGCGATC 1

Search completed: June 20, 2000, 18:25:46  
Job time: 538057 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2000, 09:51:12 ; Search time 13789.4 Seconds  
(without alignments)  
5291.175 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_196000\_214000  
Perfect score: 18001  
Sequence: 1 CTRAAAAGTATTTTAAACC.....TGGACATGCTGTGTTCTTC 18001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues  
Total number of hits satisfying chosen parameters: 156056

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result % Query

No.	Score	Match	Length	DB	ID	Description
1	93	0.5	109	30	AA243009	AA243009 zr25h02.s
2	91.6	0.5	106	37	AA703692	AA703692 aq81a10.r
3	91	0.5	107	35	AA565533	AA565533 nk42d11.s
4	90.2	0.5	103	84	B48914	B48914 RPC111-4A12
5	90	0.5	110	106	AQ386882	AQ386882 RPC111-13
6	88.6	0.5	103	108	AQ535244	AQ535244 RPC111-3
7	87.6	0.5	110	106	AQ386882	AQ386882 RPC111-13
8	87.4	0.5	106	108	AQ544957	AQ544957 CITBT-EI-
9	87.2	0.5	109	84	B17434	B17434 345K2.rVB C
10	86.8	0.5	110	38	AA897366	AA897366 am06h02.s
11	86.6	0.5	103	38	AA807640	AA807640 nx08b05.s
12	86.6	0.5	110	30	AA244245	AA244245 nc07a04.s
13	86.6	0.5	110	94	AQ003188	AQ003188 RPC111-1D
14	86.2	0.5	107	35	AA565533	AA565533 nk42b11.s
15	86.2	0.5	108	84	B65160	B65160 CIT-HSP-201
16	86	0.5	105	109	AQ637292	AQ637292 RPC111-4
17	85.4	0.5	103	94	AQ028649	AQ028649 CIT-HSP-2
18	85.4	0.5	103	108	AQ535244	AQ535244 RPC111-3
19	85	0.5	106	30	AA250812	AA250812 zs06a05.s
20	85	0.5	109	84	B17434	B17434 345K2.rVB C
21	84.4	0.5	102	36	AA654562	AA654562 nt75f10.s
22	83.6	0.5	106	105	AQ264176	AQ264176 CITBT-EI-
23	83.4	0.5	109	105	AQ265749	AQ265749 CITBT-EI-
24	83	0.5	107	35	AA583252	AA583252 nn41e04.s
25	83	0.5	109	94	AQ029690	AQ029690 RPC111-41
26	82	0.5	103	108	AQ582186	AQ582186 RPC111-4
27	82	0.5	106	63	AI991750	AI991750 wt48e01.x
28	82	0.5	106	63	AI991750	AI991750 wt48e01.x
29	81.6	0.5	107	24	H67040	H67040 yu68c01.r1
30	81.2	0.5	102	84	B48088	B48088 RPC111-4N6.
31	81.2	0.5	104	108	AQ544583	AQ544583 CITBT-EI-
32	81.2	0.5	108	35	AA594869	AA594869 no21e02.s
33	81	0.4	105	28	AA078003	AA078003 7H12D08 C
34	81	0.4	107	103	AQ240182	AQ240182 CIT-HSP-2
35	80.8	0.4	108	84	B32951	B32951 HS-1016-A1
36	80.8	0.4	109	103	AQ200347	AQ200347 RPC111-43
37	80.6	0.4	103	35	AA570438	AA570438 nk63g02.s
38	80.6	0.4	103	108	AQ534922	AQ534922 RPC111-3
39	80.6	0.4	108	84	B15423	B15423 345B10 TV C
40	80.4	0.4	106	38	AA812141	AA812141 OB48h02.s
41	80.4	0.4	109	24	N25299	N25299 yw52c09.s1
42	80.2	0.4	106	30	AA250812	AA250812 zs06a05.s
43	80.2	0.4	109	94	AQ028426	AQ028426 CIT-HSP-2
44	80.2	0.4	109	105	AQ265749	AQ265749 CITBT-EI-
45	80	0.4	97	39	AA837701	AA837701 oe06c02.s

## ALIGNMENTS

```
RESULT 1
AA243009 AA243009 109 bp mRNA EST 11-MAR-1998
LOCUS zr25h02.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens
DEFINITION cDNA clone IMAGE:664467 3' similar to contains Alu repetitive
element:contains element LTR1 repetitive element ;, mRNA sequence.
ACCESSION AA243009
VERSION AA243009.1 GI:1873869
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 109)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
JOURNAL On Dec 3, 1996 this sequence version replaced gi:1126869.
COMMENT
```

```
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1127 Std Error: 0.00
Seq primer: -41m13 fwd. ET from Amersham
High quality sequence stop: 102.
FEATURES
Location/Qualifiers
1..109
/organism="Homo sapiens"
/db_xref="GDB:5426481"
/db_xref="taxon:9606"
/clone="IMAGE:664467"
/clone_lib="Stratagene NT2 neuronal precursor 937230"
/dev_stage="Ntera-2 neuroepithelial cells"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: brain; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo df. Uninduced, exponentially growing neuroepithelial
cells (Ntera-2/ci.D1). Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT 19 a 30 c 30 g 30 t
ORIGIN
Query Match 0.5%; Score 93; DB 30; Length 109;
Best Local Similarity 90.8%; Pred. No. 0.066;
Matches 99; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
Qy 15134 GTATTTTGTAGATGAGTGGAGTTTCGCCGTGTAGCCAGGATGCTCGATCTCCTGACC 15193
Db 1 GTATTTTGTAGATGAGACGGGGTTTCACCGTGTAGCCAGGATGCTCGATCTCCTGACC 60
Qy 15194 TCGTGATCCACGGCTCGGCTCCCAAGTCTCCGATGCTGGGATTACAGGCATG 15242
Db 61 TCGTGATCCGCCACCTCGGCTCCCAAGTCTGGGATTACAGGCATG 109
RESULT 2
AA703692 106 bp mRNA EST 24-DEC-1997
LOCUS aq81a10.r1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone
DEFINITION IMAGE:1140858 5' similar to contains Alu repetitive element; , mRNA
sequence.
ACCESSION AA703692
VERSION AA703692.1 GI:2713610
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 106)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
JOURNAL On Sep 12, 1996 this sequence version replaced gi:1397630.
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
```

Seq primer: -28ml3 rev1 ET from Amersham  
High quality sequence stop: 53.

## FEATURES

## source

1. 106 Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1140858"  
/clone\_lib="Stratagene hnt neuron (#937233)"  
/dev\_stage="hnt neurons"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Differentiated, post mitotic hnt neurons. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTGTGTTTTTTTTTTT 3'."  
CTCGAGTGTGTTTTTTTTTTT 3'."

BASE COUNT 19 a 29 c 29 g 29 t

## ORIGIN

Query Match 0.5%; Score 91.6; DB 37; Length 106;  
Best Local Similarity 91.5%; Pred. No. 0.097; Mismatches 9; Indels 0; Gaps 0;  
Matches 97; Conservative 0;

QY 15137 TTTTGTAGATGAGTTCGCGTGTAGCCAGGATGTTCTCGATCTCCTGACCTCG 15196

Db 1 TTTTGTAGACGAGGTTTCCACGTTGTAGCCAGGATGTTCTCGATCTCCTGACCTCG 60

QY 15197 TGATCCAGCGCTCGGCTCCCAAGTGTGGGATTACAGGCATG 15242

Db 61 TGATGCTCCGCTCAGCTCCCAAGTGTGGGATTACAGGCATG 106

## RESULT 3

## AA565533

LOCUS AA565533 107 bp MRNA EST 08-SEP-1997

DEFINITION nk42b11.s1 NCI-CCAP\_GC2 Homo sapiens cDNA clone IMAGE:1016157 3' similar to contains Alu repetitive element;; mRNA sequence.

ACCESSION AA565533

VERSION AA565533.1 GI:2337172

KEYWORDS EST.

SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 107)

AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT On Sep 12, 1996 this sequence version replaced gi:1393355.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Stratagene, Inc., David B. Krizman,

Ph.D.

cDNA Library Arraying: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CCAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Insert Length: 1661 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 87.

## FEATURES

## source

1. 107 Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1016157"  
/clone\_lib="NCI-CCAP\_GC2"

/tissue\_type="germ cell tumor"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Bulk germ cell tumor. 5' adaptor sequence: 5' GAATTCGCGCAGAG 3' 3' adaptor sequence: 5' CTCGAGTGTGTTTTTTTTTTT 3'."  
Average insert size: 1.2 kb."  
BASE COUNT 22 a 34 c 26 g 25 t

## ORIGIN

Query Match 0.5%; Score 91; DB 35; Length 107;  
Best Local Similarity 90.7%; Pred. No. 0.11;  
Matches 97; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 15007 CTGGAGTCAGTGGTCCGATCTCGGCTCACTGCAACCTCTGCTCCCGGTTCAAGTGAT 15066

Db 1 CTGGAGTCAGTGGTCCGATCTCAATCTCACTGCAAGCTCTGCTCCCGGTTCAAGTGAT 60

QY 15067 TCTCTGCTCAGCTCCGAGTAGCTTGAGCTACAGGCACACCA 15113

Db 61 TCTGCTCAGCTCCTGCTCTGAGTAGCTGGGATTACAGGCACACCA 107

## RESULT 4

## B48914/c

LOCUS B48914 103 bp DNA GSS 08-APR-1999

DEFINITION RPC111-4A12.TP RPC1-11 Homo sapiens genomic clone RPC1-11-4A12, genomic survey sequence.

ACCESSION B48914

VERSION B48914.1 GI:2601151

KEYWORDS GSS.

SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 103)

AUTHORS Adams, M.D., Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.

TITLE Use of BAC End Sequences for Sequence-Ready Map Building

JOURNAL Unpublished (1997)

COMMENT

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdamas@tigr.org

Clones are derived from the human BAC library RPC1-11. For BAC

library availability, please contact Pieter de Jong

([pieter@dejong.med.buffalo.edu](mailto:pieter@dejong.med.buffalo.edu)). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from

Research Genetics ([info@resgen.com](http://resgen.com)). BAC end search page:

[http://www.tigr.org/tldb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html)

Seq primer: SP6

Class: BAC ends.

## FEATURES

## source

1. 103 Location/Qualifiers

/organism="Homo sapiens"

/db\_xref="GDB:7501163"

/db\_xref="taxon:9606"

/clone="RPC1-11-4A12"

/clone\_lib="RPC1-11"

/sex="Male"

/cell\_type="Lymphocytes"

/note="vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;

RPC111 Human Male BAC Library"

BASE COUNT 30 a 28 c 30 g 15 t

## ORIGIN

Query Match 0.5%; Score 90.2; DB 84; Length 103;

```
Best Local Similarity 92.2%; Pred. No. 0.14;
Matches 95; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 15133 TGTATTTTGTAGAGAGGAGTTTCGCCGTTAGCCAGAGTGTCTCGATCTCCTGAC 15192
|||||
Db 103 TGTATTTTGTAGAGAGCGGGTTTACCGTTTACCGGGATGTCPCGATCTCCTGAC 44
|||||
Qy 15193 CTCGTGATCCACCGGCTCGGCTCCCAAGTGTGGGATTAC 15235
|||||
Db 43 CTCGTGATCCGCCCTCGGCTCCCAAGTGTGGGCTTAC 1
|||||

RESULT 5
AQ386882/c
LOCUS
DEFINITION RPCI11-13414_TV RPCI-11 Homo sapiens genomic clone RPCI-11-13414,
genomic survey sequence.
ACCESSION AQ386882
VERSION AQ386882.1 GI:4357905
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
Venter,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPCI11-13414.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..110
/organism="Homo sapiens"
/db_xref="GDB:7551267"
/db_xref="taxon:9606"
/clone="RPCI-11-13414"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"
BASE COUNT 26 a 26 c 38 g 20 t
ORIGIN
Query Match 0.5%; Score 90; DB 106; Length 110;
Best Local Similarity 90.6%; Pred. No. 0.14;
Matches 96; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 15853 GGGTTTACCATTGACAGCGTGGTCTCAACTCTCGACTCAAGTGATCCACCTGCC 15912
|||||
Db 110 GGGTTTACCATTGTTGTCCAGCGTGGTCTTGAACCTCTGACCTCAAGCGATCCACCTGCC 51
|||||
Qy 15913 TTAGGCTCCCAAAATGCTGGGACTACAGCGGTGAGCCACTGCACCC 15958
|||||
Db 50 TCAGCTCCCAAAAGTACTTGGATTACAGCGGTGAGCCACTGCTCCC 5
|||||
```

```
RESULT 6
AQ353244
LOCUS
DEFINITION RPCI-11-317H22_TV RPCI-11 Homo sapiens genomic clone
RPCI-11-317H22, genomic survey sequence.
ACCESSION AQ353244
VERSION AQ353244.1 GI:4846934
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
Venter,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..103
/organism="Homo sapiens"
/db_xref="GDB:7621533"
/db_xref="taxon:9606"
/clone="RPCI-11-317H22"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"
BASE COUNT 31 a 27 c 27 g 18 t
ORIGIN
Query Match 0.5%; Score 88.6; DB 108; Length 103;
Best Local Similarity 91.3%; Pred. No. 0.22;
Matches 94; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 10405 CCAGCATTTCGAGGCTGAGCGGGGAGATCATTGAGGTCAGAGTTTGAGACCACC 10464
|||||
Db 1 CCAGCATTTCGAGGCAAGCAAGCGGAGATCATTGAGGTCAGAGTTTGAGACCACC 60
|||||
Qy 10465 TGGCAACATGTTGAACCCCTGCTCCACTAAAAATACAAAAA 10507
|||||
Db 61 TGGCAACATGTTGAACCCCGCTCTCTCTATTAATACAAAAA 103
|||||

RESULT 7
AQ386882
LOCUS
DEFINITION RPCI11-13414_TV RPCI-11 Homo sapiens genomic clone RPCI-11-13414,
genomic survey sequence.
ACCESSION AQ386882
VERSION AQ386882.1 GI:4357905
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
```



REFERENCE 8  
LOCUS AQ544957/c  
DEFINITION A0544957 106 bp DNA GSS 28-MAY-1999  
CITBI-EI-2629N2.TF CITBI-EI Homo sapiens genomic clone 2629N2,  
genomic survey sequence.  
ACCESSION A0544957  
VERSION A0544957.1 GI:4903683  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 9  
LOCUS B17434  
DEFINITION B17434 109 bp DNA GSS 04-JUN-1998  
345K2.TVB CIT978SKAL Homo sapiens genomic clone A-345K02, genomic  
survey sequence.  
ACCESSION B17434  
VERSION B17434.1 GI:2125183  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 10  
LOCUS B17434  
DEFINITION B17434 109 bp DNA GSS 04-JUN-1998  
345K2.TVB CIT978SKAL Homo sapiens genomic clone A-345K02, genomic  
survey sequence.  
ACCESSION B17434  
VERSION B17434.1 GI:2125183  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

Query Match 0.5%; Score 87.6; DB 106; Length 110;  
Best Local Similarity 87.3%; Pred. No. 0.27; Mismatches 0; Gaps 0;  
Matches 96; Conservative 0; Indels 0; Gaps 0;  
QY 10375 GCGCGGCGCTGGTTCAGCTGATCCAGCAGCTTGGCAGGCTGAGCGGCGCAGA 10434  
DB 1 GCGCGGAGCAGTGGCTCAGCGCTGATCCAGTACTTTGGGAGCTGAGCAGGTGGA 60  
QY 10435 TCACTTGAGTCAGGAGTTTGAGACAGCAGCTGCCACATCGTGAACCC 10484  
DB 61 TCGCTTGAGTCAGGAGTTTCAGACAGCAGCTGGACACATGTTGAACCC 110  
RESULT 8  
LOCUS AQ544957/c  
DEFINITION A0544957 106 bp DNA GSS 28-MAY-1999  
CITBI-EI-2629N2.TF CITBI-EI Homo sapiens genomic clone 2629N2,  
genomic survey sequence.  
ACCESSION A0544957  
VERSION A0544957.1 GI:4903683  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

Query Match 0.5%; Score 87.4; DB 108; Length 106;  
Best Local Similarity 89.5%; Pred. No. 0.29; Mismatches 0; Gaps 0;  
Matches 94; Conservative 0; Indels 0; Gaps 0;  
QY 15155 TTTCGCGGTGTTAGCCAGGATGCTCTGATCTCTGACCTCGTATCCACCGGCTCGGC 15214  
DB 106 TTTCACCGTGTAGCCAGAAATGCTTGTATCTCTGACCTAAATGATCTGCCGCTCGGC 47  
QY 15215 CTCCCAAAGTCTGGGATTTACAGGCTGCGCCACCGCTGGCC 15259  
DB 46 CTCCCAAAGTCTGGGATTTACAGGCTGCGCCCTCCAGCGCCGCGCC 2  
RESULT 9  
LOCUS B17434  
DEFINITION B17434 109 bp DNA GSS 04-JUN-1998  
345K2.TVB CIT978SKAL Homo sapiens genomic clone A-345K02, genomic  
survey sequence.  
ACCESSION B17434  
VERSION B17434.1 GI:2125183  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

Query Match 0.5%; Score 87.4; DB 108; Length 106;  
Best Local Similarity 89.5%; Pred. No. 0.29; Mismatches 0; Gaps 0;  
Matches 94; Conservative 0; Indels 0; Gaps 0;  
QY 15155 TTTCGCGGTGTTAGCCAGGATGCTCTGATCTCTGACCTCGTATCCACCGGCTCGGC 15214  
DB 106 TTTCACCGTGTAGCCAGAAATGCTTGTATCTCTGACCTAAATGATCTGCCGCTCGGC 47  
QY 15215 CTCCCAAAGTCTGGGATTTACAGGCTGCGCCACCGCTGGCC 15259  
DB 46 CTCCCAAAGTCTGGGATTTACAGGCTGCGCCCTCCAGCGCCGCGCC 2  
RESULT 9  
LOCUS B17434  
DEFINITION B17434 109 bp DNA GSS 04-JUN-1998  
345K2.TVB CIT978SKAL Homo sapiens genomic clone A-345K02, genomic  
survey sequence.  
ACCESSION B17434  
VERSION B17434.1 GI:2125183  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

Query Match 0.5%; Score 87.4; DB 108; Length 106;  
Best Local Similarity 89.5%; Pred. No. 0.29; Mismatches 0; Gaps 0;  
Matches 94; Conservative 0; Indels 0; Gaps 0;  
QY 15155 TTTCGCGGTGTTAGCCAGGATGCTCTGATCTCTGACCTCGTATCCACCGGCTCGGC 15214  
DB 106 TTTCACCGTGTAGCCAGAAATGCTTGTATCTCTGACCTAAATGATCTGCCGCTCGGC 47  
QY 15215 CTCCCAAAGTCTGGGATTTACAGGCTGCGCCACCGCTGGCC 15259  
DB 46 CTCCCAAAGTCTGGGATTTACAGGCTGCGCCCTCCAGCGCCGCGCC 2  
RESULT 9  
LOCUS B17434  
DEFINITION B17434 109 bp DNA GSS 04-JUN-1998  
345K2.TVB CIT978SKAL Homo sapiens genomic clone A-345K02, genomic  
survey sequence.  
ACCESSION B17434  
VERSION B17434.1 GI:2125183  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

Query Match 0.5%; Score 87.4; DB 108; Length 106;  
Best Local Similarity 89.5%; Pred. No. 0.29; Mismatches 0; Gaps 0;  
Matches 94; Conservative 0; Indels 0; Gaps 0;  
QY 15155 TTTCGCGGTGTTAGCCAGGATGCTCTGATCTCTGACCTCGTATCCACCGGCTCGGC 15214  
DB 106 TTTCACCGTGTAGCCAGAAATGCTTGTATCTCTGACCTAAATGATCTGCCGCTCGGC 47  
QY 15215 CTCCCAAAGTCTGGGATTTACAGGCTGCGCCACCGCTGGCC 15259  
DB 46 CTCCCAAAGTCTGGGATTTACAGGCTGCGCCCTCCAGCGCCGCGCC 2  
RESULT 9  
LOCUS B17434  
DEFINITION B17434 109 bp DNA GSS 04-JUN-1998  
345K2.TVB CIT978SKAL Homo sapiens genomic clone A-345K02, genomic  
survey sequence.  
ACCESSION B17434  
VERSION B17434.1 GI:2125183  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

BASE COUNT      24 a 30 c 31 g 24 t
ORIGIN

Query Match      0.5%; Score 87.2; DB 84; Length 109;
Best Local Similarity 88.0%; Pred. No. 0.3;
Matches 95; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1650 GCCTCATACCTGTATATCCAGCAGCTTTGGAGACTGGAGTGGGATCATCTGAGGTCA 1709
|||||
Db 2 GCCTCATACCTATATATCTAGCAGCTTTGGAGGCTGATGGCGGATCACTGAGGTCTG 61
|||||

Qy 1710 GGAGTCAAAACAGCCCTGGCCACATGGTGAACACCATCTCTACTA 1757
|||||
Db 62 GGAGTTCGAGACCGCTGGCCACCATGGTGAACACCGCTCTCACTA 109
|||||

RESULT 10
AA897366
LOCUS
DEFINITION
am06h02.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:1466067 3' similar to contains Alu repetitive element;; mRNA
sequence.
ACCESSION
AA897366
VERSION
AA897366.1 GI:3033986
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 110)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
On Jan 19, 1998 this sequence version replaced gi:2150764.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 834 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 63.
FEATURES
Location/Qualifiers
1..110
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1466067"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pTT3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP-GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
22 a 27 c 29 g 32 t

BASE COUNT      22 a 27 c 29 g 32 t
ORIGIN

Query Match      0.5%; Score 86.8; DB 39; Length 110;
Best Local Similarity 88.7%; Pred. No. 0.34;
Matches 94; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 15137 TTTTGTAGAGATGGAGTTTCGGCGGTGTAGCCAGGATGGTCTCGATCTCCTGACCTCG 15196
|||||

```

```

Db 5 TTTTGTAGAGATGGGTTTTCACCGTGTAAACCAGGATGGTCTCAATCTCTGGACCTCA 64
|||||

Qy 15197 TGATCCACCGCGCTCGGCTCCCAAAGTCTGGGATTACAGGCATG 15242
|||||
Db 65 TGATCCGCGCCACCTCGGCTCCCAAAGTCTGGGATTATAGGCGTG 110
|||||

RESULT 11
AA807640
LOCUS
DEFINITION
nx08b05.s1 NCI-CGAP_GC3 Homo sapiens cDNA clone IMAGE:1255473 3'
similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION
AA807640
VERSION
AA807640.1 GI:2877108
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 103)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
On Jan 19, 1998 this sequence version replaced gi:2151346.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 774 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 87.
FEATURES
Location/Qualifiers
1..103
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1255473"
/clone_lib="NCI-CGAP_GC3"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pTT3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled
germ cell tumors, and was then primed with a Not I -
oligo(UT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pTT3
vector. Library is not normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo."
19 a 27 c 30 g 27 t

BASE COUNT      19 a 27 c 30 g 27 t
ORIGIN

Query Match      0.5%; Score 86.6; DB 38; Length 103;
Best Local Similarity 91.1%; Pred. No. 0.37;
Matches 92; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 15142 AGTAGAGATGGAGTTTCGGCGGTGTAGCCAGGATGGTCTCGATCTCCTGACCTCGTATC 15201
|||||
Db 2 AGTAGAGATGGGTTTTCACCGTGTAGCCAGGATGGTCTCGATCTCCTGACCTGTGATC 61
|||||

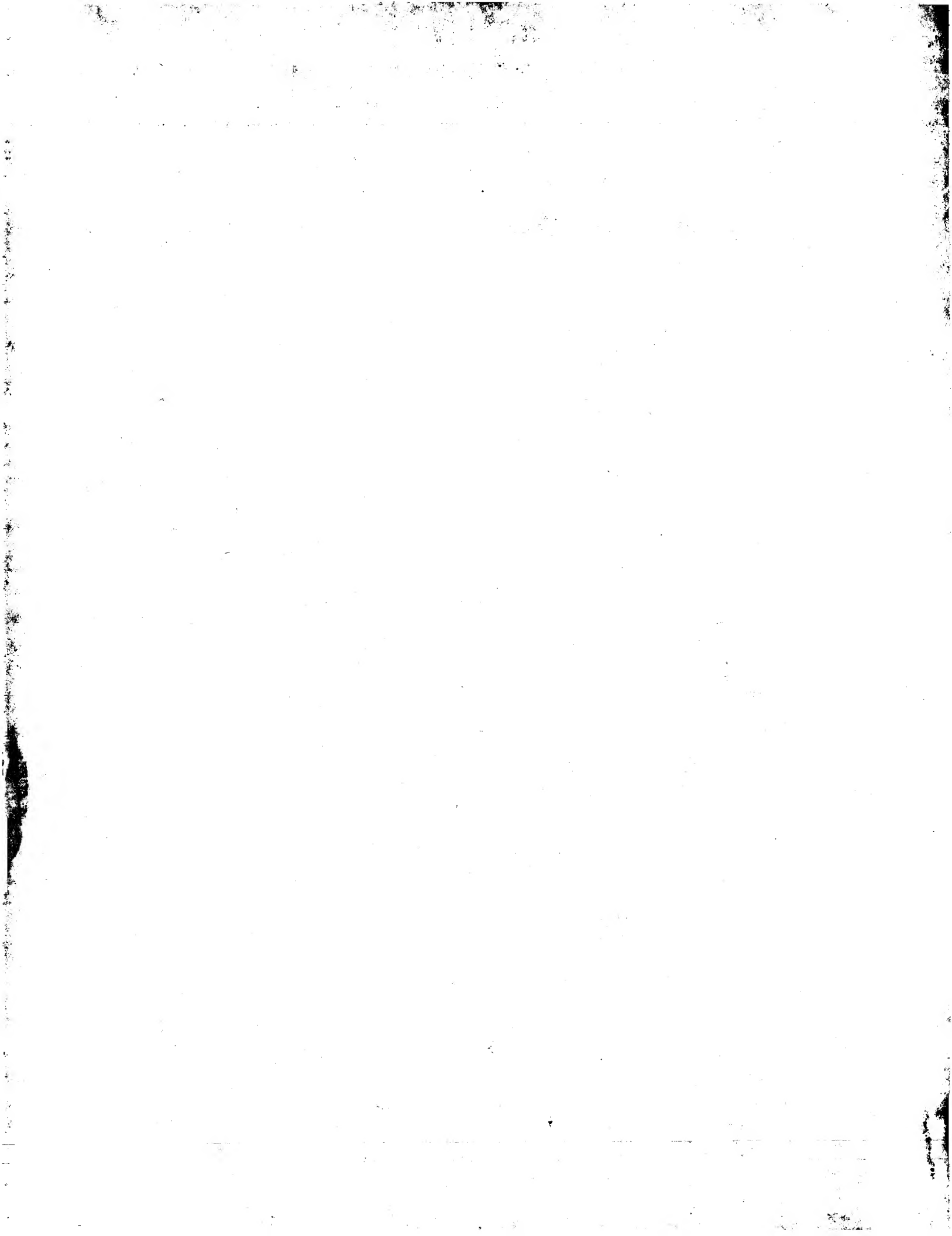
Qy 15202 CACCGGCTCGGCTCCCAAAGTCTGGGATTACAGGCATG 15242
|||||
Db 62 CGCTCACCTCGGCTCCCAAAGTCTGGGATTACAGGTTG 102
|||||

```

RESULT 12  
AA244245  
LOCUS  
DEFINITION nc07a04.s1 NCI\_CGAP\_Prl Homo sapiens cDNA clone IMAGE:1007406  
similar to contains Alu repetitive element;; mRNA sequence.  
ACCESSION AA244245  
VERSION AA244245.1 GI:1875104  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 110)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 24, 1995 this sequence version replaced gi:634306.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
CDNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Seq primer: -4lm13 fwd. ET from Amersham  
High quality sequence stop: 90.  
FEATURES  
source  
1..110  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1007406"  
/clone.lib="NCI\_CGAP\_Prl"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"  
/note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from  
5,000-10,000 microdissected, histologically normal  
prostate epithelial cells. Double-stranded cDNA was  
ligated to EcoRI adaptors, 5 cycles of PCR applied to the  
cDNA with an adaptor-specific primer, and the resulting  
PCR product subcloned into pAMP10 by the UDG-cloning  
method (Life Technologies). Average insert size is 600  
bp. NOTE: Not directionally cloned. This library was  
constructed by David Krizman."  
BASE COUNT 17 a 26 c 28 g 38 t 1 others  
ORIGIN  
Query Match 0.5%; Score 86.6; DB 30; Length 110;  
Best Local Similarity 86.4%; Pred. No. 0.36;  
Matches 95; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
QY 15671 TTTTGTGAGATGCTCATCTGTCTACCCAGGCTGGAGTACAGTACATGATCT 15730  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 1 TTTTGTGAGATGAGTCTGTATCTGTGCTCCAGGCTGGAGTGGAGTGGCAGTCT 60  
QY 15731 TGGCTCATGCAACCTCGCCCTCTGGGTTCAGTGTCTCTCTGACTCA 15780  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 61 TGGCTCACTGCAACCTCTGCTCTCTGGGTTCAGAGATTTCTTCTGCTCTCA 110  
RESULT 13  
AQ003188  
LOCUS  
DEFINITION RPCI11-1D10.TPN RPCI-11 Homo sapiens genomic clone RPCI-11-1D10,  
genomic survey sequence.  
ACCESSION AQ003188  
VERSION AQ003188.1 GI:3030392  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 110)  
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,  
Golden,K., Barry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and  
Venter,J.C.  
TITLE Use of BAC End Sequences for Sequence-Ready Map Building (1998)  
JOURNAL Unpublished (1998)  
COMMENT Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdamads@tigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from  
Research Genetics ([info@resgen.com](http://info@resgen.com)). BAC end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: SP6  
Class: BAC ends.  
FEATURES  
Location/Qualifiers  
1..110  
/organism="Homo sapiens"  
/db\_xref="GDB:7500081"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-1D10"  
/clone.lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPCI11 Human Male BAC Library"  
BASE COUNT 22 a 27 c 26 g 35 t  
ORIGIN  
Query Match 0.5%; Score 86.6; DB 94; Length 110;  
Best Local Similarity 87.2%; Pred. No. 0.36;  
Matches 95; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
QY 15129 TTTTGTATTTTACTAGAGATGGAGTTTCGCGGTGTAGCCAGGATGCTCGATCTCC 15188  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 2 TTTTGTATTTTACTAGATACAGGGTTTACCATGTTGCGCCAGGATGCTCCGATCTCT 61  
QY 15189 TGACCTCGTGATCCACCGGCTCGGCTCCCAAGTCTGGGATTACAG 15237  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 62 TGACCTCATGATCCACCTGCCCGACGCTCCCAAGTCTGGGATTACAG 110  
RESULT 14  
AA565533/c  
LOCUS  
DEFINITION AA565533 107 bp mRNA EST 08-SEP-1997  
nk42b11.s1 NCI\_CGAP\_GC2 Homo sapiens cDNA clone IMAGE:1016157 3'  
similar to contains Alu repetitive element;; mRNA sequence.  
ACCESSION AA565533  
VERSION AA565533.1 GI:2337172  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 107)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),



---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2000, 18:07:04 ; Search time 599.42 Seconds  
(without alignments)  
3903.542 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_196000\_214000  
Perfect score: 18001  
Sequence: 1 CTAAGAAAGTATTTAAAC.....TGGACATGCTGTCTTCCTTC 18001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 374504

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78.4	0.4	105	4	US-08-481-658B-65
2	78.4	0.4	105	4	US-08-477-504A-65
3	78.4	0.4	105	4	US-08-486-756A-65
4	78.4	0.4	105	4	US-08-485-862B-65
5	78.4	0.4	105	5	US-08-787-739-65
6	72.8	0.4	105	4	US-08-481-658B-65
7	72.8	0.4	105	4	US-08-477-504A-65
8	72.8	0.4	105	4	US-08-486-756A-65
9	72.8	0.4	105	4	US-08-485-862B-65
10	72.8	0.4	105	5	US-08-787-739-65
11	64	0.4	92	1	US-08-222-177A-430
12	63.8	0.4	84	3	US-08-454-557C-91
13	63.8	0.4	84	4	US-08-340-426D-91
14	63.8	0.4	84	4	US-08-450-673C-91
15	63.8	0.4	84	6	PCT-US95-17111A-91
16	60.4	0.3	78	3	US-08-454-557C-70
17	60.4	0.3	78	4	US-08-340-426D-70
18	60.4	0.3	78	4	US-08-450-673C-70
19	60.4	0.3	78	6	PCT-US95-17111A-70
20	60.2	0.3	98	1	US-08-088-658-42
21	60.2	0.3	98	4	US-08-471-907A-42
22	59.2	0.3	85	3	US-08-454-557C-92
23	59.2	0.3	85	4	US-08-340-426D-92
24	59.2	0.3	85	4	US-08-450-673C-92
25	59.2	0.3	85	6	PCT-US95-17111A-92
26	59	0.3	60	1	US-08-222-177A-244
27	58.2	0.3	84	3	US-08-454-557C-91

c 28	58.2	0.3	84	4	US-08-340-426D-91	Sequence 91, Appl
c 29	58.2	0.3	84	4	US-08-450-673C-91	Sequence 91, Appl
c 30	58.2	0.3	84	6	PCT-US95-17111A-91	Sequence 91, Appl
c 31	57.8	0.3	78	3	US-08-454-557C-70	Sequence 70, Appl
c 32	57.8	0.3	78	4	US-08-340-426D-70	Sequence 70, Appl
c 33	57.8	0.3	78	4	US-08-450-673C-70	Sequence 70, Appl
c 34	57.8	0.3	78	6	PCT-US95-17111A-70	Sequence 70, Appl
c 35	55.8	0.3	85	3	US-08-332-766A-44	Sequence 44, Appl
c 36	55.4	0.3	76	3	US-08-454-557C-69	Sequence 69, Appl
c 37	55.4	0.3	76	4	US-08-340-426D-69	Sequence 69, Appl
c 38	55.4	0.3	76	4	US-08-450-673C-69	Sequence 69, Appl
c 39	55.4	0.3	76	6	PCT-US95-17111A-69	Sequence 69, Appl
c 40	53.6	0.3	60	3	US-08-454-557C-60	Sequence 60, Appl
c 41	53.6	0.3	60	4	US-08-340-426D-60	Sequence 60, Appl
c 42	53.6	0.3	60	4	US-08-450-673C-60	Sequence 60, Appl
c 43	53.6	0.3	60	6	PCT-US95-17111A-60	Sequence 60, Appl
c 44	53.6	0.3	91	1	US-08-222-177A-166	Sequence 166, App
c 45	52.6	0.3	60	3	US-08-454-557C-57	Sequence 57, Appl

ALIGNMENTS

RESULT 1  
US-08-481-658B-65  
; Sequence 65, Application US/08481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,658B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.3E  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-481-658B-65

Query Match 0.4%; Score 78.4; DB 4; Length 105;  
Best Local Similarity 84.6%; Pred. No. 6.9e-09;

Matches 88; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 15129 TTTTGTATTTTGTAGATGAGATGAGTTTCGCCGTGTTAGCCAGATGGTCTCGATCTCC 15188

Db 2 TTTTACATCTTTAGTAGACACAGGGTTTACCATATTTGCCAGGCTGCTCTCAAACTCC 61

QY 15189 TGACCTCGTGATCCACCGGCTCGGCCCTCCCAAAAGTGTGGAT 15232

Db 62 TGACCTGTGTATCCACCAAGCTCGGCCCTCCCAAAAGTGTGGAT 105

## RESULT 2

US-08-477-504A-65

; Sequence 65, Application US/08477504A

; Patent No. 5972353

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

; CITY: Tiburon

; STATE: California

; COUNTRY: USA

; ZIP: 94920

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/477,504A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/260,190

; FILING DATE: 15-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lauder, Leona L.

; REGISTRATION NUMBER: 30,863

; REFERENCE/DOCKET NUMBER: D-0021.3D

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-435-2034

; TELEFAX: 415-435-0727

; INFORMATION FOR SEQ ID NO: 65:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 105 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-477-504A-65

Query Match 0.4%; Score 78.4; DB 4; Length 105;

Best Local Similarity 84.6%; Pred. No. 6.9e-09;

Matches 88; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 15129 TTTTGTATTTTGTAGATGAGATGAGTTTCGCCGTGTTAGCCAGATGGTCTCGATCTCC 15188

Db 2 TTTTACATCTTTAGTAGACACAGGGTTTACCATATTTGCCAGGCTGCTCTCAAACTCC 61

QY 15189 TGACCTCGTGATCCACCGGCTCGGCCCTCCCAAAAGTGTGGAT 15232

Db 62 TGACCTGTGTATCCACCAAGCTCGGCCCTCCCAAAAGTGTGGAT 105

## RESULT 3

US-08-486-756A-65

; Sequence 65, Application US/08486756A

; Patent No. 5981711

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

; CITY: Tiburon

; STATE: California

; COUNTRY: USA

; ZIP: 94920

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/486,756A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/260,190

; FILING DATE: 15-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lauder, Leona L.

; REGISTRATION NUMBER: 30,863

; REFERENCE/DOCKET NUMBER: D-0021.3C

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-435-2034

; TELEFAX: 415-435-0727

; INFORMATION FOR SEQ ID NO: 65:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 105 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-486-756A-65

Query Match 0.4%; Score 78.4; DB 4; Length 105;

Best Local Similarity 84.6%; Pred. No. 6.9e-09;

Matches 88; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 15129 TTTTGTATTTTGTAGATGAGATGAGTTTCGCCGTGTTAGCCAGATGGTCTCGATCTCC 15188

Db 2 TTTTACATCTTTAGTAGACACAGGGTTTACCATATTTGCCAGGCTGCTCTCAAACTCC 61

QY 15189 TGACCTCGTGATCCACCGGCTCGGCCCTCCCAAAAGTGTGGAT 15232

Db 62 TGACCTGTGTATCCACCAAGCTCGGCCCTCCCAAAAGTGTGGAT 105

## RESULT 4

US-08-485-862B-65

; Sequence 65, Application US/08485862B

; Patent No. 5989838

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court



CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485.862B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-485-862B-65

Query Match 0.4%; Score 78.4; DB 4; Length 105;  
Best Local Similarity 84.6%; Pred. No. 6.9e-09;  
Matches 88; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 15129 TTTTGTATTTTGTAGAGATGGAGTTTCGCCGTTAGCCAGGATGCTCTCGATCTCC 15188  
||||| |||||||| ||||| || ||||||| ||||| ||||| |||||  
Db 2 TTTTACATCTTTAGTAGACAGGTTTTCACCATATTGGCCAGGCTGCTCTCAAACTCC 61

Qy 15189 TCACCTCGTGATCCACCGGCTCGGCCCTCCCAAAAGTCTGGGAT 15232  
||||| |||||||| |||||||| |||||||| |||||||| |||||||| ||||||||  
Db 62 TGACCTTGTGATCCACCGGCTCGGCCCTCCCAAAAGTCTGGGAT 105

RESULT 5  
US-08-787-739-65  
; Sequence 65, Application US/08787739  
; Patent No. 6027887  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 96  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 369 Pine Street, Suite 610  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/787,739  
FILING DATE: 24-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,049  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/486,756  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/481,658  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,862  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,863  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/487,077  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.4  
TELEPHONE: 415-981-2034  
TELEFAX: 415-981-0332  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-787-739-65

Query Match 0.4%; Score 78.4; DB 5; Length 105;  
Best Local Similarity 84.6%; Pred. No. 6.9e-09;  
Matches 88; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 15129 TTTTGTATTTTGTAGAGATGGAGTTTCGCCGTTAGCCAGGATGCTCTCGATCTCC 15188  
||||| |||||||| ||||| || ||||||| ||||| ||||| |||||  
Db 2 TTTTACATCTTTAGTAGACAGGTTTTCACCATATTGGCCAGGCTGCTCTCAAACTCC 61

Qy 15189 TCACCTCGTGATCCACCGGCTCGGCCCTCCCAAAAGTCTGGGAT 15232  
||||| |||||||| |||||||| |||||||| |||||||| |||||||| ||||||||  
Db 62 TGACCTTGTGATCCACCGGCTCGGCCCTCCCAAAAGTCTGGGAT 105

RESULT 6  
US-08-481-658B-65/c  
; Sequence 65, Application US/08481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:



MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-486-756A-65

Query Match 0.4%; Score 72.8; DB 4; Length 105;  
Best Local Similarity 86.8%; Pred. No. 1.2e-07;  
Matches 92; Conservative 0; Mismatches 12; Indels 2; Gaps 1;

QY 10402 ATCCGAGCATTGGCAGGCTGAGGGGCGAGATCATTGAGTGAGGAGTTGAGACCA 10461  
|||||  
DB 105 ATCCGAGCATTGGGAGGCCGAGGCTGGTGATCAC--AAGGTGAGGAGTTGAGAGCA 48  
|||||  
QY 10462 GCCTGCCAACATGTGTAACCCCTGTCTCCACTAAAAATACAAAA 10507  
|||||  
DB 47 GCCTGCCAATATGTTGAACCCCTGTCTCTACTAAGATGTAATAA 2  
|||||

RESULT 9  
US-08-485-862B-65/c  
; Sequence 65, Application US/08485862B  
; Patent No. 5989838  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,862B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/477,504  
; FILING DATE: 07-JUN-1995  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.3D  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-485-862B-65

Query Match 0.4%; Score 72.8; DB 4; Length 105;  
Best Local Similarity 86.8%; Pred. No. 1.2e-07;  
Matches 92; Conservative 0; Mismatches 12; Indels 2; Gaps 1;

QY 10402 ATCCGAGCATTGGCAGGCTGAGGGGCGAGATCATTGAGTGAGGAGTTGAGACCA 10461  
|||||  
DB 105 ATCCGAGCATTGGGAGGCCGAGGCTGGTGATCAC--AAGGTGAGGAGTTGAGAGCA 48  
|||||  
QY 10462 GCCTGCCAACATGTGTAACCCCTGTCTCCACTAAAAATACAAAA 10507  
|||||  
DB 47 GCCTGCCAATATGTTGAACCCCTGTCTCTACTAAGATGTAATAA 2  
|||||  
RESULT 10  
US-08-787-739-65/c  
; Sequence 65, Application US/08787739  
; Patent No. 6027887  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 96  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 369 Pine Street, Suite 610  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/787,739  
; FILING DATE: 24-JAN-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,049  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/486,756  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/477,504  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/481,658  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,862  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,863  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/487,077  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-981-2034  
; TELEFAX: 415-981-0332  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-787-739-65



```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 14-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; US-08-340-426D-91

Query Match 0.4%; Score 63.8; DB 4; Length 84;
Best Local Similarity 85.5%; Pred. No. 1.1e-05;
Matches 71; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 15160 CCGTGTACCCAGGATGGTCTCGATCTCTGACCTGCTGATCCACGGGCTCGGCTCCC 15219
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1 CCATGTTTCATCAGGCTGGTGTGCAACTCTGACCTGCTGATCGGCCCTCAGCCTCCC 60

QY 15220 AAGTGTCTGGGATTACAGCGTG 15242
||||||| ||||| ||||| |||||
Db 61 AAGTGTCTGGGATTACAGCGTG 83

RESULT 15
PCT-US95-17111A-91
; Sequence 91, Application PC/TUS9517111A
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and
; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/17111A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/340,426
; FILING DATE: 14-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; PCT-US95-17111A-91

Query Match 0.4%; Score 63.8; DB 6; Length 84;
Best Local Similarity 85.5%; Pred. No. 1.1e-05;
Matches 71; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 15160 CCGTGTACCCAGGATGGTCTCGATCTCTGACCTGCTGATCCACGGGCTCGGCTCCC 15219
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1 CCATGTTTCATCAGGCTGGTGTGCAACTCTGACCTGCTGATCGGCCCTCAGCCTCCC 60

QY 15220 AAGTGTCTGGGATTACAGCGTG 15242
||||||| ||||| ||||| |||||
Db 61 AAGTGTCTGGGATTACAGCGTG 83

RESULT 14
US-08-450-673C-91
; Sequence 91, Application US/08450673C
; Patent No. 5948888
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,673C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
```

```
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; US-08-450-673C-91

Query Match 0.4%; Score 63.8; DB 4; Length 84;
Best Local Similarity 85.5%; Pred. No. 1.1e-05;
Matches 71; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 15160 CCGTGTACCCAGGATGGTCTCGATCTCTGACCTGCTGATCCACGGGCTCGGCTCCC 15219
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1 CCATGTTTCATCAGGCTGGTGTGCAACTCTGACCTGCTGATCGGCCCTCAGCCTCCC 60

QY 15220 AAGTGTCTGGGATTACAGCGTG 15242
||||||| ||||| ||||| |||||
Db 61 AAGTGTCTGGGATTACAGCGTG 83

RESULT 15
PCT-US95-17111A-91
; Sequence 91, Application PC/TUS9517111A
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and
; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/17111A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/340,426
; FILING DATE: 14-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; PCT-US95-17111A-91

Query Match 0.4%; Score 63.8; DB 6; Length 84;
Best Local Similarity 85.5%; Pred. No. 1.1e-05;
Matches 71; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 15160 CCGTGTACCCAGGATGGTCTCGATCTCTGACCTGCTGATCCACGGGCTCGGCTCCC 15219
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1 CCATGTTTCATCAGGCTGGTGTGCAACTCTGACCTGCTGATCGGCCCTCAGCCTCCC 60

QY 15220 AAGTGTCTGGGATTACAGCGTG 15242
||||||| ||||| ||||| |||||
```

Db 61 AACTGTCTGGATTACAAGCGTG 83

Search completed: June 20, 2000, 18:07:26  
Job time: 537176 sec







```

MEDLINE      97415756
REFERENCE    2 (bases 1 to 108)
AUTHORS      Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE        Direct Submission
JOURNAL      Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The
              Children's Hospital of Philadelphia, 1004F Abramson Research
              Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA
FEATURES
  source      1..108
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="TscAlu2"
              1..108
  repeat_region
              /note="scAlu"
              /rpt_family="Alu"
              /rpt_type=dispersed
  BASE COUNT 23 a 39 c 30 g 16 t
  ORIGIN
              0.4%; Score 89.6; DB 11; Length 108;
              Best Local Similarity 95.8%; Pred. No. 0.00037;
              Matches 92; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 12730 GTAGACGGGTTTACCATGTTAGCCAGGATGCTCTCGATCTCTGACCTCGTGATCC 12789
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 97 GTAGACGGGTTTACCTTGTGTTAGCCAGGATGCTCTCGATCTCTGACCTCGTGATCC 38

Qy 12790 GCCCACTGAGCCTCCCAAGTGTGGGATTACAGG 12825
          |||| |||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37 GCCCGCTCGGCTCCCAAGTGTGGGATTACAGG 2

RESULT      5
HSLDLRD1
LOCUS       HSLDLRD1 108 bp DNA PRI 20-MAY-1992
DEFINITION Human LDL-receptor mutated gene with intron 12 deletion junction.
ACCESSION  X05249
VERSION     X05249.1 GI:34335
KEYWORDS    Alu repetitive sequence; low density lipoprotein receptor.
SOURCE      human.
ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 108)
AUTHORS     Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
              Williamson,R. and Humphries,S.
TITLE       Unequal crossing-over between two alu-repetitive DNA sequences in
              the low-density-lipoprotein-receptor gene. A possible mechanism for
              the defect in a patient with familial hypercholesterolaemia
              Eur. J. Biochem. 164 (1), 77-81 (1987)
JOURNAL     87161901
MEDLINE
COMMENT     See X05248 for corresponding normal gene sequence
              In the defective LDL-receptor gene the deletion occurred between two
              alu-repetitive sequences, that are in the same direction, the
              deletion eliminates exons 13 and 14 and changes the reading frame
              of the resulting spliced mRNA.
              Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES
  source      1..108
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /cell_type="blood leukocytes from a patient with familial"
              1..108
  BASE COUNT 28 a 20 c 40 g 20 t
  ORIGIN
              Query Match 0.4%; Score 87.8; DB 10; Length 108;
              Best Local Similarity 88.8%; Pred. No. 0.00067;
              Matches 95; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 16884 TTGGCTACTGCAACCTCTGCTCTCTGCTTCAACGAATTCCTCCTCAGCCTTCCAA 16943
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 107 TCGCCTACCAACACCTCTGCTCTCTGCTTCAACCAACATTTCTCGCTCAGCCTCCCGA 61

Qy 16944 GTAGCTGGGATTACAGGACCCACCTACCAACGCTGGCTAAATTTTGT 16990
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 GTAGCTGGGATTACAGGACCTGCCACCACGCTGGCTAAATTTTGT 108

RESULT      6
HSLDLRD2/c
LOCUS       HSLDLRD2 108 bp DNA PRI 20-MAY-1992
DEFINITION Human LDL-receptor mutated gene with intron 14 deletion junction.
ACCESSION  X05251
VERSION     X05251.1 GI:34336
KEYWORDS    Alu repetitive sequence; low density lipoprotein receptor.
SOURCE      human.
ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 108)
AUTHORS     Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
              Williamson,R. and Humphries,S.
TITLE       Unequal crossing-over between two alu-repetitive DNA sequences in
              the low-density-lipoprotein-receptor gene. A possible mechanism for
              the defect in a patient with familial hypercholesterolaemia
              Eur. J. Biochem. 164 (1), 77-81 (1987)
JOURNAL     87161901
MEDLINE
COMMENT     See X05250 for corresponding normal gene sequence
              In the defective LDL-receptor gene the deletion occurred between two
              alu-repetitive sequences, that are in the same direction, the
              deletion eliminates exons 13 and 14 and changes the reading frame
              of the resulting spliced mRNA.
              Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES
  source      1..108
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /cell_type="blood leukocytes from a patient with familial"
              1..108
  BASE COUNT 28 a 20 c 40 g 20 t
  ORIGIN
              Query Match 0.4%; Score 87.8; DB 10; Length 108;
              Best Local Similarity 88.8%; Pred. No. 0.00067;
              Matches 95; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 16884 TTGGCTACTGCAACCTCTGCTCTCTGCTTCAACGAATTCCTCCTCAGCCTTCCAA 16943
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 107 TCGCCTACCAACACCTCTGCTCTCTGCTTCAACCAACATTTCTCGCTCAGCCTCCCGA 48

Qy 16944 GTAGCTGGGATTACAGGACCCACCTACCAACGCTGGCTAAATTTTGT 16990
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 47 GTAGCTGGGATTACAGGACCTGCCACCACGCTGGCTAAATTTTGT 1

RESULT      7
HUMALCE221/c
LOCUS       HUMALCE221 103 bp ss-RNA PRI 15-APR-1994
DEFINITION Human carcinoma cell-derived Alu RNA transcript, clone CE221.
ACCESSION  M87896
VERSION     M87896.1 GI:174874
KEYWORDS    Alu repeat.
SOURCE      Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 103)
AUTHORS     Sinnott,D., Richer,C., Deragon,J.-M. and Labuda,D.

```





```
|||||
Db 6 CCTGTAATCCAGCTACACGGAGCTAAGCAGGAGAGTTCGTTGAACCGGGAGCGG 65
Qy 7333 AGTTGCAATGAGCAAGATCATGCACTGCACCTCCAG 7370
Db 66 AGTTGCAATGAGCGGAGATGCTGCCATTGCACCTCCAG 103

RESULT 14
HUMALCE43/c HUMALCE43 110 bp ss-RNA PRI 15-APR-1994
LOCUS Human carcinoma cell-derived Alu RNA transcript, clone CE43..
DEFINITION M87900.1 GI:174876
ACCESSION M87900.1
VERSION 1
KEYWORDS Alu repeat.
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 110)
AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
post-transcriptional selection of master sequences
JOURNAL J. Mol. Biol. (1992) in press
FEATURES
    source
        1..110
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /cell_line="Wt9a2D1"
            /dev_stage="embryo"
            /sex="male"
            /tissue_type="carcinoma"
BASE COUNT 27 a 31 c 34 g 18 t
ORIGIN

Query Match 0.3%; Score 78.8; DB 9; Length 110;
Best Local Similarity 84.0%; Pred. No. 0.014;
Matches 89; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 12741 GTTTCACCATGTTAGCCAGGATGGTCTCGATCTCTGACCTCGTGATCCGCCACCTGAG 12800
Db 110 GTTTCGTCATGTGACCGAGGTGGTCTTGAACACTAGCTGCGCAATCTCTCGCTTGG 51
Qy 12801 CCTCCAAAGTGGGATACAGGTGTGAGCCGCGCCGCC 12846
Db 50 CCTCCAAAGTGGCGGATGTAGTGTGAGCCGCGCCGCC 5

RESULT 15
HSLDL12
LOCUS HSLDL12 108 bp DNA PRI 20-MAY-1992
DEFINITION Human LDL-receptor gene intron 12 fragment (normal gene) LbL - low
density lipoprotein.
ACCESSION X05248
VERSION X05248.1 GI:34334
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor;
repetitive sequence.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
Williamson,R. and Humphries,S.
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT see X05249 for deletion junction
DATA kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES
    Location/Qualifiers
```

```
source 1..108
misc_feature /db_xref="taxon:9606"
intron complement(<1..65)
note="Alu repeat"
1..108
/note="intron XII fragment"
BASE COUNT 21 a 38 c 20 g 29 t
ORIGIN

Query Match 0.3%; Score 76.8; DB 10; Length 108;
Best Local Similarity 83.7%; Pred. No. 0.027;
Matches 87; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 20565 TCGGCTCACAGCAACCTCCACCTCTCTGGGTTCAAGTGATTTCTCTGCCTCAGCCTCCTGA 20624
Db 2 TCGGCTCACAGCAACCTCTCTGGGTTCAAGTGATTTCTCTGCCTCAGCCTCCTGA 61
Qy 20625 GTAGCTGGGATTACAGACGCTGTGCACACACACCTGGCTAAATTTT 20668
Db 62 GTAGCTGGGATTACAGACGCTGTGCACACACCGCCGGCTGATTTT 105

Search completed: June 21, 2000, 05:17:36
Job time: 578340 sec
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2000, 18:25:46 ; Search time 490.16 Seconds  
(without alignments)  
12417.216 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_213000\_237326  
Perfect score: 24327  
Sequence: 1 TAGAGTTAAATGTGAAAAAT.....TGTGTGTGTGTGTGTGTGTGTG 24327

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 433070

Minimum DB seq length: 10

Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	67.6	0.3	108	1	X12095	Human biallelic po
2	67.2	0.3	100	1	T24892	Human gene signatu
3	64.4	0.3	108	1	T26828	Human gene signatu
4	64	0.3	108	1	T25009	Human gene signatu
5	63.4	0.3	108	1	T25009	Human gene signatu
6	62.6	0.3	91	1	T25854	Human gene signatu
7	60.8	0.2	100	1	T24892	Human gene signatu
8	60	0.2	103	1	T26213	Human gene signatu
9	59.6	0.2	97	1	T26728	Human gene signatu
10	58.4	0.2	103	1	T26213	Human gene signatu
11	58	0.2	103	1	T20927	Human gene signatu
12	56.4	0.2	107	1	T20373	Human gene signatu
13	56.6	0.2	108	1	T26828	Human gene signatu
14	55.6	0.2	100	1	X12087	Human biallelic po
15	55.6	0.2	100	1	X12085	Human gene signatu
16	54.8	0.2	87	1	T21566	Human gene signatu
17	54.8	0.2	93	1	T22572	Human gene signatu
18	55	0.2	95	1	T23131	Human gene signatu
19	54.4	0.2	93	1	T25868	Human gene signatu
20	54.2	0.2	93	1	T24259	Human gene signatu
21	54.2	0.2	93	1	T24259	Human gene signatu
22	54.2	0.2	95	1	T23131	Human gene signatu
23	54.4	0.2	100	1	X12086	Human biallelic po
24	54.4	0.2	106	1	Q95210	Simple tandem repe
25	53.6	0.2	69	1	Q29016	Probe to internal
26	53	0.2	91	1	T25854	Human gene signatu
27	52.6	0.2	110	1	T26288	Human gene signatu
28	51.4	0.2	82	1	T65723	Repeat sequence fr
29	51.6	0.2	93	1	T22572	Human gene signatu
30	51.4	0.2	109	1	T23895	Human gene signatu
31	51	0.2	70	1	N60231	Normal chromosome
32	50.4	0.2	99	1	T20931	Human gene signatu
33	50	0.2	69	1	T24175	Human gene signatu
34	50	0.2	75	1	T22841	Human gene signatu

C 35	50	0.2	81	1	T24093	Human gene signatu
C 36	48.8	0.2	62	1	Q34053	Microsatellite seq
C 37	49	0.2	65	1	T24893	Human gene signatu
C 38	48.6	0.2	99	1	T20931	Human gene signatu
C 39	48.4	0.2	99	1	T23728	Human gene signatu
C 40	48	0.2	91	1	T26410	Human gene signatu
C 41	48.2	0.2	101	1	T24667	Human gene signatu
C 42	48.4	0.2	105	1	T23942	Human gene signatu
C 43	48.6	0.2	106	1	V11611	Homo sapiens adult
C 44	47.4	0.2	65	1	T25588	Human gene signatu
C 45	47.6	0.2	82	1	T25468	Human gene signatu

ALIGNMENTS

RESULT 1

X12095  
ID X12095 standard; DNA; 108 BP.  
AC X12095;  
DT 30-MAR-1999 (first entry)  
DE Human biallelic polymorphic DNA fragment TIGR-A003M18a.  
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;  
KW detection; phenotypic typing; characteristic; infection; hereditary;  
KW autoimmune disease; cancer; inflammation; drug; therapy; medication;  
KW treatment; marker; ss.  
OS Homo sapiens.  
PN W09820165-A2.  
PD 14-MAY-1998.  
PF 05-NOV-1997; U20313.  
PR 06-NOV-1996; US-030455.  
PA (WHEED ) WHITEHEAD INST BIOMEDICAL RES.  
PI Hudson T, Lander ES, Wang D;  
DR WPI; 98-286374/25.  
PT New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease  
PS Claim 1; Page 219; 310pp; English.  
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
CC markers which have been isolated using the primers represented in  
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, Ehlers-Danlos  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.  
SQ Sequence 108 BP; 19 A; 23 C; 28 G; 37 T;

Query Match 0.3%; Score 67.6; DB 1; Length 108;  
Best Local Similarity 82.6%; Pred. No. 0.21;  
Matches 76; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

QY 16988 TGTATTTCTGTAGACAGGGGTTTTCACCATGTTGGCCAGGCTGCTCTCGAATCTCTGAC 17047

Db 1 TGTCTTTTGTAGAGATGAGGTTTCTCTGTGGCCAGGATGCTCTCGAATCTCTGAC 60

QY 17048 CTCAGGTGATCCCAATCTTGGCCTCCCAA 17079

Db 61 TTCAAGTGATCCGCTGCTGCTTGGCCTCCCAA 92

```
RESULT 2
ID T24892/c
AC T24892;
DE T24892 standard; cDNA to mRNA; 100 BP.
DT 05-NOV-1996 (first entry)
DE Human gene signature HUMGS06998.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI: 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1720; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 100 BP; 28 A; 22 C; 25 G; 22 T;

Query Match 0.3%; Score 67.2; DB 1; Length 100;
Best Local Similarity 78.8%; Pred. No. 0.24;
Matches 78; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 14066 TATTATTATTTTCAGACGAGTCTCACTGTCTCACCACGAGCTGAGTGCGCGCATC 14125
DB 100 TTTGTTGTTTTCACACAGAGTCTCACTGTCTGTCACCCAGCNGAGTGAANGTGCAATC 41

QY 14126 TCGGCTCACTGCAACCTCCGCTCCCGGTTCAAGTGAT 14164
DB 40 TCAGCTNATTGCAAAATCTGCTCCCGAGTTCACGGAT 2

RESULT 3
ID T26828
AC T26828 standard; cDNA to mRNA; 108 BP.
DT 14-NOV-1996 (first entry)
DE Human gene signature HUMGS09078.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI: 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
```

```
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 2182; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 108 BP; 18 A; 33 C; 23 G; 28 T;

Query Match 0.3%; Score 64.4; DB 1; Length 108;
Best Local Similarity 89.5%; Pred. No. 0.56;
Matches 68; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 3995 ATCTCTGACCTCGGATCCGCCGCTCCGCTCCCAAGTGGGATTACAGCGTG 4054
DB 2 ATCTCTGACCTCGGATCCGCCGCTCCGCTCCCAAGTGGGATTACAGCGTG 61

QY 4055 AGCCACCGTGCCTCGC 4070
DB 62 AGCCACCGTGCCTCGC 77

RESULT 4
ID T25009/c
AC T25009 standard; cDNA to mRNA; 108 BP.
DT 07-NOV-1996 (first entry)
DE Human gene signature HUMGS07131.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI: 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1748; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
```





PS Claim 1; Page 1720; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 100 BP; 28 A; 22 C; 25 G; 22 T;

Query Match 0.2%; Score 60.8; DB 1; Length 100;  
Best Local Similarity 74.7%; Pred. No. 1.7;  
Matches 74; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 7311 ATCGTTGAACCCAGAGGTGGAGTTCATGACCAAGATCATGCCACTGCATCCAG 7370  
Db 2 ATCGTTGAACCTGGAGGAGCAATTTGCAATNAGCTGAGATTGCACCTTGCACCTCCG 61

QY 7371 CCTGGGCACAGAGTAAGACTCCGTTTCAAAAACAAAAA 7409  
Db 62 CCTGGGTGACAGAGTGACACTCTGTGTTGAAACAAACAA 100

RESULT 8  
T26213  
ID T26213 standard; cDNA to mRNA; 103 BP.  
AC T26213;  
DE Human gene signature HUMGS08452.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN W09514772-A1.  
PD 01-JUN-1995.  
PR 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI; 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1; Page 2029; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 103 BP; 33 A; 21 C; 25 G; 23 T;

Query Match 0.2%; Score 60; DB 1; Length 103;  
Best Local Similarity 75.0%; Pred. No. 2.2;  
Matches 75; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 7311 ATCGTTGAACCCAGAGGTGGAGTTCATGACCAAGATCATGCCACTGCATCCAG 7370  
Db 2 ATCATCTTGAGTCCAGAGAGTTGGTGTGTACAGTGAGCTATGATGGCACCACCTGCCTCCAG 61

QY 7371 CCTGGGCACAGAGTAAGACTCCGTTTCAAAAACAAAAA 7410  
Db 62 CCTGGGCACAGAGTAAGACTGTCTTTAAGAAAAAAA 101

RESULT 9  
T26728  
ID T26728 standard; cDNA to mRNA; 97 BP.  
AC T26728;  
DE Human gene signature HUMGS08978.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN W09514772-A1.  
PD 01-JUN-1995.  
PR 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI; 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1; Page 2158; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 97 BP; 19 A; 27 C; 20 G; 28 T;

Query Match 0.2%; Score 59.6; DB 1; Length 97;  
Best Local Similarity 81.0%; Pred. No. 2.5;  
Matches 68; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 14297 GATCTACCCACCTTGGCTCCCAAGAGTCTGGGATTACAGGATGAGCCACTGCGCCAG 14356  
Db 1 GATCTGCCCACTGCTCCAGAGTCTGGGATTACAGGATGAGCCACTGCGCCAG 60

QY 14357 CCTGCTGTCTTTTATTTTATA 14380  
Db 61 NCTGTACTAAGTCTTTTTTTTATA 84

RESULT 10  
T26213/c  
ID T26213 standard; cDNA to mRNA; 103 BP.  
AC T26213;  
DE Human gene signature HUMGS08452.

KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN WO9514772-AL.  
PD 01-JUN-1995.

PF 11-NOV-1994; J01916.

PR 12-NOV-1993; JP-355504.

PA (MATS/) MATSUBARA K.

PA (OKUB/) OKUBO K.

PI Matsubara K, Okubo K;

PI WPI; 95-206931/27.

DR Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues

PS Claim 1; Page 2039; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.

SQ Sequence 103 BP; 33 A; 21 C; 25 G; 23 T;

Query Match 0.2%; Score 58.4; DB 1; Length 103;

Best Local Similarity 74.0%; Pred. No. 3.5;

Matches 74; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 14065 TTATTTATTTTTCAGCGGAGTCTACTCTGTCACCGCTGAGTGCAGTGGCCGAT 14124

DB 101 TTTTTTTCTTAAAGACATGTTCTTACTCTGTGGCGCTGAGTGCAGTGGCCCAT 42

QY 14125 CTGGCTCACTGCAACCTCGCTCCCGGTTCAAGTGAT 14164

DB 41 CATAGCTCACTGTAAACCAACCACTCTGGACATCAAGTGAT 2

RESULT 11

ID T20927 standard; cDNA to mRNA; 103 BP.

AC T20927;

DE Human gene signature HUMGS0180.

KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;

KW cell typing; abnormal cell function; ss.

OS Homo sapiens.

PN WO9514772-AL.

PD 01-JUN-1995.

PF 11-NOV-1994; J01916.

PR 12-NOV-1993; JP-355504.

PA (MATS/) MATSUBARA K.

PA (OKUB/) OKUBO K.

PI Matsubara K, Okubo K;

PI WPI; 95-206931/27.

DR Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues

PS Claim 1; Page 758-759; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be

CC determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.

SQ Sequence 103 BP; 22 A; 27 C; 31 T;

Query Match 0.2%; Score 58; DB 1; Length 103;

Best Local Similarity 73.0%; Pred. No. 4;

Matches 73; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 20601 GATTCCTCCCTCAGCCCTCCTCAGTAGCTGGGATTACAGACCTGTGTACACACCTGG 20660

DB 1 GATCCTCCCACTTCCACCTCCCAAGTAGCTGTGGCTGTGTGCCACCATGTCCAG 60

QY 20661 CTAATTTTGTATTTTAGTAGACAGACAGGGTTTCACCGTG 20700

DB 61 CTGATTTTNGTATTTTNTAGTAGGACAGTATTTCTCCATG 100

RESULT 12

T20373/c

ID T20373 standard; cDNA to mRNA; 107 BP.

AC T20373;

DE Human gene signature HUMGS01525.

KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;

KW human; cloning; mapping; non-biased library; diagnosis; detection;

KW cell typing; abnormal cell function; ss.

OS Homo sapiens.

PN WO9514772-AL.

PD 01-JUN-1995.

PF 11-NOV-1994; J01916.

PR 12-NOV-1993; JP-355504.

PA (MATS/) MATSUBARA K.

PA (OKUB/) OKUBO K.

PI Matsubara K, Okubo K;

PI WPI; 95-206931/27.

DR Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human

PT tissues

PS Claim 1; Page 623; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in T19001-T26837 and which is able to hybridise to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost

CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library

CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be

CC determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.

SQ Sequence 107 BP; 26 A; 29 C; 17 G; 29 T;

Query Match

Best Local Similarity 72.6%; Pred. No. 6.5;

Matches 69; Conservative 0; Mismatches 26; Indels 0; Gaps 0;



CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, familial colonic polyposis, Ehlers-Danlos  
CC haemorrhagic telangiectasia, acute intermittent porphyria,  
CC syndrome, osteogenesis imperfecta, cancer, diseases of the nervous  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases. 22 A; 25 C; 22 G; 30 T;  
SQ Sequence 100 BP; 22 A; 25 C; 22 G; 30 T;

Query Match 0.2%; Score 55.6; DB 1; Length 100;  
Best Local Similarity 72.9%; Pred. No. 8.4;  
Matches 70; Conservative 1; Mismatches 25; Indels 0; Gaps 0;  
QY 14682 TCTTGCTGTGTGCTCAGCGCTGGTGTGAACCTCTGGGCTTCAAGGATCCTCTGCCTTG 14741  
Db 4 TCTTGCTGTGTGCTCAGCGCTGGTGTGAACCTCTGGGCTTCAAGGATCCTCTGCCTTG 63  
QY 14742 GTCTCACAATGCTGGGATGACAGACATGACAC 14777  
Db 64 GCCTCCYAAAGTGCCAGGATTATAGGTGTGATGCAC 99

Search completed: June 21, 2000, 07:57:11  
Job time: 586742 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2000, 09:51:36 ; Search time 7160.88 Seconds  
(without alignments)  
13769.643 Million cell updates/sec

Title: US-08-852-495C-2\_COPY\_213000\_237326  
Perfect score: 24327  
Sequence: 1 TAGAGTTAAATGTGAAAAAT.....TGTGTGTGTGTGTGTGTG 24327

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 156056

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*

- 1: em\_est1:\*
- 2: em\_est2:\*
- 3: em\_est3:\*
- 4: em\_est4:\*
- 5: em\_est5:\*
- 6: em\_est6:\*
- 7: em\_est7:\*
- 8: em\_est8:\*
- 9: em\_est9:\*
- 10: em\_est10:\*
- 11: em\_est11:\*
- 12: em\_est12:\*
- 13: em\_est13:\*
- 14: em\_est14:\*
- 15: em\_est15:\*
- 16: em\_est16:\*
- 17: em\_est17:\*
- 18: em\_est18:\*
- 19: em\_est19:\*
- 20: gb\_est1:\*
- 21: gb\_est2:\*
- 22: gb\_est3:\*
- 23: gb\_est4:\*
- 24: gb\_est5:\*
- 25: gb\_est6:\*
- 26: gb\_est7:\*
- 27: gb\_est8:\*
- 28: gb\_est9:\*
- 29: gb\_est10:\*
- 30: gb\_est11:\*
- 31: gb\_est12:\*
- 32: gb\_est13:\*
- 33: gb\_est14:\*
- 34: gb\_est15:\*
- 35: gb\_est16:\*
- 36: gb\_est17:\*
- 37: gb\_est18:\*
- 38: gb\_est19:\*
- 39: gb\_est20:\*
- 40: gb\_est21:\*
- 41: gb\_est22:\*
- 42: gb\_est23:\*
- 43: gb\_est24:\*
- 44: gb\_est25:\*

- 45: gb\_est26:\*
- 46: gb\_est27:\*
- 47: gb\_est28:\*
- 48: gb\_est29:\*
- 49: gb\_est30:\*
- 50: gb\_est31:\*
- 51: gb\_est32:\*
- 52: em\_est20:\*
- 53: em\_est21:\*
- 54: em\_est22:\*
- 55: em\_est23:\*
- 56: em\_est24:\*
- 57: em\_est25:\*
- 58: em\_est26:\*
- 59: gb\_est33:\*
- 60: gb\_est34:\*
- 61: gb\_est35:\*
- 62: gb\_est36:\*
- 63: gb\_est37:\*
- 64: gb\_est38:\*
- 65: em\_est27:\*
- 66: em\_est28:\*
- 67: em\_est29:\*
- 68: em\_est30:\*
- 69: gb\_est39:\*
- 70: gb\_est40:\*
- 71: gb\_est41:\*
- 72: gb\_est42:\*
- 73: gb\_est43:\*
- 74: gb\_est44:\*
- 75: em\_est31:\*
- 76: em\_est32:\*
- 77: em\_est33:\*
- 78: em\_est34:\*
- 79: gb\_est45:\*
- 80: gb\_est46:\*
- 81: gb\_est47:\*
- 82: gb\_gss1:\*
- 83: gb\_gss2:\*
- 84: gb\_gss3:\*
- 85: gb\_gss4:\*
- 86: em\_gss1:\*
- 87: em\_gss2:\*
- 88: em\_gss3:\*
- 89: em\_gss4:\*
- 90: gb\_gss5:\*
- 91: gb\_gss6:\*
- 92: gb\_gss7:\*
- 93: gb\_gss8:\*
- 94: gb\_gss9:\*
- 95: em\_gss5:\*
- 96: em\_gss6:\*
- 97: em\_gss7:\*
- 98: em\_gss8:\*
- 99: em\_gss9:\*
- 100: em\_gss10:\*
- 101: em\_gss11:\*
- 102: gb\_gss10:\*
- 103: gb\_gss11:\*
- 104: em\_gss12:\*
- 105: gb\_gss12:\*
- 106: gb\_gss13:\*
- 107: gb\_gss14:\*
- 108: gb\_gss15:\*
- 109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	96.4	0.4	106	37	AA703692	ag81a10.r
2	94.8	0.4	109	30	AA243009	zr25h02.s
3	92.4	0.4	103	38	AA807640	nx08b05.s
4	92.6	0.4	108	84	B65160	CIT-HSP-201
5	92.4	0.4	110	39	AA897366	am06h02.s
6	91.4	0.4	109	94	AQ028426	CIT-HSP-2
7	91	0.4	108	84	B65160	CIT-HSP-201
8	90.4	0.4	110	30	AA244245	nc07a04.s
9	89.8	0.4	101	39	AA835205	ak64h01.s
10	88.4	0.4	106	105	AQ264176	CITBT-EI-
11	87.8	0.4	103	84	B48914	RPC111-4A12
12	87.8	0.4	107	35	AA565533	nk42b11.s
13	87	0.4	103	30	AA228795	nc14e07.s
14	86.8	0.4	103	108	AQ582186	RPCI-11-4
15	87	0.4	103	108	AQ584425	RPCI-11-4
16	86.6	0.4	103	108	AQ535244	RPCI-11-3
17	86.8	0.4	106	38	AA812141	ob48h02.s
18	86.4	0.4	105	30	AA218889	zgl5d04.s
19	86.4	0.4	107	39	AA828124	od71a07.s
20	86.6	0.4	109	94	AQ028426	CIT-HSP-2
21	86.6	0.4	110	30	AA244245	nc07a04.s
22	85.6	0.4	102	30	AA228656	nc19f09.s
23	85.8	0.4	105	28	AA078003	7H12D08 C
24	85.8	0.4	105	105	AQ282107	RPCI11-94
25	85.2	0.4	106	94	AQ062963	CIT-HSP-2
26	85.2	0.4	109	22	H11143	YM08C06.r1
27	84.8	0.3	104	29	AA129957	zn86h04.r
28	84.8	0.3	104	29	AA129957	zn86h04.r
29	84.8	0.3	107	33	AA385808	EST99495
30	85	0.3	109	84	B17434	345K2.TVB C
31	85	0.3	109	105	AQ265749	CITBT-EI-
32	84.6	0.3	107	103	AQ240182	CIT-HSP-2
33	83.8	0.3	108	94	AQ014433	CIT-HSP-2
34	83.4	0.3	101	39	AA835205	ak64h01.s
35	83.4	0.3	102	94	AQ004934	CIT-HSP-2
36	83.6	0.3	109	24	N25299	YW52C09.s1
37	83	0.3	101	33	AA381369	EST94442
38	83.2	0.3	104	108	AQ544583	CITBT-EI-
39	83.2	0.3	106	106	AQ418993	RPCI-11-1
40	83.2	0.3	107	24	N23686	YW46a02.s1
41	83.4	0.3	110	94	AQ003188	RPCI11-1D
42	82.6	0.3	106	108	AQ544957	CITBT-EI-
43	82.8	0.3	110	29	AA177157	nc02g07.s
44	82	0.3	98	24	H67549	YU68f10.s1
45	82.2	0.3	104	105	AQ268072	RPCI11-73

## ALIGNMENTS

```
RESULT 1
AA703692
LOCUS      AA703692      106 bp      mRNA      EST      24-DEC-1997
DEFINITION ag81a10.r1 Strata gene hnt neuron (#937233) Homo sapiens cDNA clone
            IMAGE:1140858 5' similar to contains Alu repetitive element; mRNA
            sequence.
ACCESSION  AA703692
VERSION    AA703692.1 GI:2713610
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 106)
AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geissel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
            WashU-NCI human EST Project
            Unpublished (1997)
TITLE     WashU-NCI human EST Project
JOURNAL
COMMENT   On Sep 12, 1996 this sequence version replaced gi:1397630.
```

```
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28ml3 revl ET from Amersham
High quality sequence stop: 53.
Location/Qualifiers
1..106
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1140858"
/clone_lib="Stratagene hnt neuron (#937233)"
/dev_stage="hnt neurons"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dt.
Differentially, post mitotic hnt neurons. Average insert
size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
CTCAGAGTTTTTTTTTTT 3'
BASE COUNT 19 a 29 c 29 g 29 t
ORIGIN
Query Match 0.4%; Score 96.4; DB 37; Length 106;
Best Local Similarity 94.3%; Pred. No. 0.14;
Matches 100; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 12724 TTTTGTAGAGACGGGTTTCCACCATGTTAGCCAGGATGTCGATCTCTGACCTCG 12783
Db 1 TTTTGTAGAGACGGGTTTCCACCATGTTAGCCAGGATGTCGATCTCTGACCTCG 60
QY 12784 TGATCGCCGCTCAGGCTCCCAAAGTCTGGGATTACAGGTGTG 12829
Db 61 TGATCGCCGCTCAGGCTCCCAAAGTCTGGGATTACAGGCGTG 106
RESULT 2
AA243009 109 bp mRNA EST 11-MAR-1998
LOCUS      AA243009
DEFINITION zr25h02.s1 Strata gene NT2 neuronal precursor 937230 Homo sapiens
            cDNA clone IMAGE:664467 3' similar to contains Alu repetitive
            element; contains element LTR1 repetitive element ; mRNA sequence.
ACCESSION  AA243009
VERSION    AA243009.1 GI:1873869
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 109)
AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geissel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
            WashU-NCI human EST Project
            Unpublished (1997)
TITLE     WashU-NCI human EST Project
JOURNAL
COMMENT   On Dec 3, 1996 this sequence version replaced gi:1126869.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1127 Std Error: 0.00
Seq primer: -41ml3 fwd. ET from Amersham
High quality sequence stop: 102.
```



## FEATURES

Location/Qualifiers  
1. .103

/organism="Homo sapiens"  
/db\_xref="GDB:5426481"  
/db\_xref="taxon:9606"  
/clone="IMAGE:664467"  
/clone\_lib="Stratagene NT2 neuronal precursor 937230"  
/tissue\_type="neuroepithelial cells"  
/dev\_stage="Ntera-2 neuroepithelial cells"  
/lab\_host="SOIR (kanamycin resistant)"  
/note="Organ: brain; Vector: plasmid SK-; Site: 1;  
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:  
oligo dt. Uninduced, exponentially growing neuroepithelial  
cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGGAG  
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3"

19 a 30 c 30 g 30 t

BASE COUNT  
ORIGIN

Query Match 0.4%; Score 94.8; DB 30; Length 109;  
Best Local Similarity 93.4%; Pred. No. 0.2;  
Matches 99; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 12724 TTTTGTAGACAGCGGGTTTCCACCATGTTAGCCAGGATGCTCGATCTCCTGACCTCG 12783  
|||||  
Db 4 TTTTGTAGACAGCGGGTTTCCACCATGTTAGCCAGGATGCTCGATCTCCTGACCTCG 63  
|||||

QY 12784 TGATCGGCCACCTGAGCTCCCAAGTCTGGGATTCAGAGTG 12829  
|||||  
Db 64 TGATCGGCCACCTGCGCTCCCAAGTCTGGGATTCAGAGCGTG 109  
|||||

RESULT 3  
AA807640

LOCUS AA807640 103 bp mRNA EST 05-MAR-1998  
DEFINITION nx08b05.s1 NCI-CGAP\_GC3 Homo sapiens cDNA clone IMAGE:1255473 3'  
similar to contains Alu repetitive element;; mRNA sequence.

ACCESSION AA807640  
VERSION AA807640.1 GI:2877108  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 103)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2151346.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)

Insert Length: 774 Std Error: 0.00  
Seq primer: -40m13 fwd. Et from Amersham  
High quality sequence stop: 87.

Location/Qualifiers  
1. .103

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1255473"  
/clone\_lib="NCI-CGAP\_GC3"  
/tissue\_type="pooled germ cell tumors"

FEATURES  
source

/lab\_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from 3 pooled  
germ cell tumors, and was then primed with a Not I -  
oligo(dT) primer. Double-stranded cDNA was ligated to Eco  
RI adaptors (Pharmacia), digested with Not I and cloned  
into the Not I and Eco RI sites of the modified pT73  
vector. Library is not normalized. Library was  
constructed by Bento Soares and M. Fatima Bonaldo."

19 a 27 c 30 g 27 t

BASE COUNT  
ORIGIN

Query Match 0.4%; Score 92.4; DB 38; Length 103;  
Best Local Similarity 94.1%; Pred. No. 0.38;  
Matches 96; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 12729 AGTAGACGGGGTTTCCACCATGTTAGCCAGGATGCTCGATCTCCTGACCTCGTGATC 12788  
|||||  
Db 2 AGTAGAGATGGGGTTTCCACCATGTTAGCCAGGATGCTCGATCTCCTGACCTTGATC 61  
|||||

QY 12789 CGCCACCTGAGCTCCCAAGTCTGGGATTCAGGTGTA 12830  
|||||  
Db 62 CGCTCACCTCGGCTCCCAAGTCTGGGATTCAGGTGTA 103  
|||||

## RESULT 4

B65160 108 bp DNA GSS 21-JUN-1998  
LOCUS CIT-HSP-2017G2.TRB CIT-HSP Homo sapiens genomic clone 2017G2,  
genomic survey sequence.

ACCESSION B65160  
VERSION B65160.1 GI:2639138  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 108)  
AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,  
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,  
Simon,M. and Venter,J.C.  
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map  
Building  
JOURNAL Unpublished (1997)  
COMMENT Other\_GSSs: CIT-HSP-2017G2.TFB  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are available from Research Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC  
end search page:  
[http://www.tigr.org/tdb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html)  
Seq primer: M13 Reverse  
Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
1. .108  
/organism="Homo sapiens"  
/db\_xref="GDB:7043860"  
/db\_xref="taxon:9606"  
/clone="2017G2"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
HindIII"

26 a 27 c 34 g 21 t

BASE COUNT  
ORIGIN



```

REFERENCE
AUTHORS 1 (bases 1 to 108)
        Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
        Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
        Simon,M. and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
        Building
JOURNAL Unpublished (1997)
COMMENT Other GSSs: CIT-HSP-2017G2.TFB
        Contact: Mark Adams
        Department of Eukaryotic Genomics
        The Institute for Genomic Research
        9712 Medical Center Dr., Rockville, MD 20850, USA
        Tel: 301 838 0200
        Fax: 301 838 0208
        Email: mdadams@tigr.org
        Clones are available from Research Genetics (info@resgen.com). BAC
        end search page:
        http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES
source 1..108
        /organism="Homo sapiens"
        /db_xref="GDB:7043860"
        /db_xref="taxon:9606"
        /clone="2017G2"
        /clone_lib="CIT-HSP"
        /sex="Male"
        /cell_type="Sperm"
        /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
        HindIII"
BASE COUNT 26 a 27 c 34 g 21 t
ORIGIN
Query Match 0.4%; Score 91; DB 84; Length 108;
Best Local Similarity 90.7%; Pred. No. 0.52;
Matches 97; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 20547 GAGTCAGTGGGATGATCTCGGCTACACCAACCTCCACCTCTGGTTCAAGTGATTC 20606
        |||||
Db 107 GTGTGAGTGGGATGATCTGTGGCTACTGCAACCTCCACCTCCCGGGTTCAAGAGATTC 48
        |||||
QY 20607 CTGCTCAGCTCTCTGAGTCTGAGTCTGGGATTCAGACAGCTGTGTGCACCA 20653
        |||||
Db 47 CTGCTCAGCTCTCTGAGTCTGAGTCTGGGATTCAGGCGCATGCCACCA 1
        |||||
RESULT 8
LOCUS AA244245 110 bp mRNA EST 20-AUG-1997
DEFINITION nc07a04.s1 NCI_CGAP.Pr1 Homo sapiens cDNA clone IMAGE:1007406
        similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION AA244245
VERSION AA244245.1 GI:1875104
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
        Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 110)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
        Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 24, 1995 this sequence version replaced gi:634306.
        Contact: Robert Strausberg, Ph.D.
        Tel: (301) 496-1550
        Email: Robert_Strausberg@nih.gov
        Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuauqui,
        M.D., Michael Emmert-Buck, M.D., Ph.D.
        cDNA Library Preparation: David B. Krizman, Ph.D.
        cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -41ml3 fwd. ET from Amersham
High quality sequence stop: 90.
FEATURES
source 1..110
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:1007406"
        /clone_lib="NCI_CGAP_Pr1"
        /sex="Male"
        /dev_stage="45 years old"
        /lab_host="DH10B"
        /note="Vector: PAMP10; Site_1: NotI; Site_2: EcoRI; 1st
        strand cDNA was primed with oligo(dT)17 on 50 ng of
        DNase-treated, total cellular RNA obtained from
        5,000-10,000 microdissected, histologically normal
        prostate epithelial cells. Double-stranded cDNA was
        ligated to EcoRI adaptors. 5 cycles of PCR applied to the
        cDNA with an adaptor-specific primer, and the resulting
        PCR product subcloned into PAMP10 by the UDG-cloning
        method (Life Technologies). Average insert size is 600
        bp. NOTE: Not directionally cloned. This library was
        constructed by David Krizman."
BASE COUNT 17 a 26 c 28 g 38 t 1 others
ORIGIN
Query Match 0.4%; Score 90.4; DB 30; Length 110;
Best Local Similarity 89.0%; Pred. No. 0.6;
Matches 97; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 20507 TTTTGTGAGATGGAGTCTTGTCTGTGCTGCCAGCTGGAGTGGCATGATCTC 20566
        |||||
Db 2 TTTTGTGAGATGGAGTCTTGTCTGTGCTGCCAGCTGGAGTGGCATGATCTT 61
        |||||
QY 20567 GGTCTACACACCTCCACCTCTGGTTCAAGTGATTCCTCGCTCA 20615
        |||||
Db 62 GGCTCACTGCAACCTCTGCTCTCTGGTTCAAGAGATTCCTTCTGCTCA 110
        |||||
RESULT 9
LOCUS AA835205 101 bp mRNA EST 23-FEB-1998
DEFINITION ak64h01.s1 Barstead pancreas HPLRB1 Homo sapiens cDNA clone
        IMAGE:1412689 3' similar to contains Alu repetitive
        element;contains element KER repetitive element ;, mRNA sequence.
ACCESSION AA835205
VERSION AA835205.1 GI:2908933
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
        Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 101)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,
        Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
        Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
        Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
        WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT On Nov 29, 1993 this sequence version replaced gi:636191.
        Contact: Wilson RK
        Washington University School of Medicine
        4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
        Tel: 314 286 1800
        Fax: 314 286 1810
        Email: est@watson.wustl.edu
        This clone is available royalty-free through LLNL; contact the
        IMAGE Consortium (info@image.llnl.gov) for further information.

```

```
FEATURES
  source
Seq primer: -40ml3 fwd. ET from Amersham.
  1. .101
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="IMAGE:1412689"
  /clone_lib="Barstead pancreas HPLRB1"
  /sex="female"
  /dev_stage="adult, 34 years"
  /lab_host="DH10B"
  /note="Organ: pancreas; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Site_1: EcoRI; Site_2: NotI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5',
TGTTACGAATCTGAAGTGGCAGCGCGCCCTTTTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(AATCGGATCCTTG), digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead."
  14 a 36 c 27 g 24 t
BASE COUNT
ORIGIN

Query Match 0.4%; Score 89.8; DB 39; Length 101;
Best Local Similarity 93.1%; Pred. No. 0.72;
Matches 94; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 14076 TGAGACGAGTCTCACTGTGACCCAGCGTGGAGTGGCGGCGGCTCACT 14135
Db 1 TGAGACGAGTCTCACTGTGCGCCAGCGTGGAGTGGCGGCTGATCTCGGCTCACT 60

Qy 14136 GCAACCTCCGCTCCGCGTTCAAAGTGAATCTCTCGCTCA 14176
Db 61 GCAAGCTCCGCTCCGCGTTCACGCCATCTCTCGCTCA 101

RESULT 10
LOCUS AQ264176/c 106 bp DNA GSS 27-OCT-1998
DEFINITION CITBI-EL-2509A2.TF CITBI-EL Homo sapiens genomic clone 2509A2,
genomic survey sequence.
ACCESSION AQ264176
VERSION AQ264176.1 GI:3792743
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 106)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
JOURNAL Unpublished (1998)
COMMENT Other GSSs: CITBI-EL-2509A2.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
  1. .106
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="2509A2"

FEATURES
  source
Location/Qualifiers
  1. .106
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="2509A2"
  /sex="female"
  /dev_stage="adult, 34 years"
  /lab_host="DH10B"
  /note="Organ: pancreas; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Site_1: EcoRI; Site_2: NotI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5',
TGTTACGAATCTGAAGTGGCAGCGCGCCCTTTTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(AATCGGATCCTTG), digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead."
  14 a 36 c 27 g 24 t
BASE COUNT
ORIGIN
```

```
FEATURES
  source
Location/Qualifiers
  1. .103
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="RPCI-11-4A12"
  /clone_lib="RPCI-11"
  /sex="Male"
  /cell_type="Lymphocytes"
  /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC library"
  30 a 28 c 30 g 15 t
BASE COUNT
ORIGIN

Query Match 0.4%; Score 87.8; DB 84; Length 103;
Best Local Similarity 92.9%; Pred. No. 1.2;
Matches 92; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

BASE COUNT
ORIGIN
  25 a 30 c 34 g 17 t
  /clone_lib="CITBI-EL"
  /sex="male"
  /cell_type="sperm"
  /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"

Query Match 0.4%; Score 88.4; DB 105; Length 106;
Best Local Similarity 89.6%; Pred. No. 0.99;
Matches 95; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 12739 GGGTTTCACCATGTAGCAGGATGCTCGATCTCTGACTCTGATCGCGCCACTG 12798
Db 106 GGGTTTCACCATGTAGCAGGACGGTCTTGATCTCTGACTCTGATCCACCGCCTC 47

Qy 12799 AGCCTCCCAAGTCTGGGATTACAGTGTGAGCCACGCGCCCGG 12844
Db 46 GGTCTCCCAAGTCTGGGATTACAGCGGTGAGACTCTGCGCCCGG 1

RESULT 11
LOCUS B48914/c 103 bp DNA GSS 08-APR-1999
DEFINITION RPCI11-4A12.TP RPCI-11 Homo sapiens genomic clone RPCI-11-4A12,
genomic survey sequence.
ACCESSION B48914
VERSION B48914.1 GI:2601151
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 103)
AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
Venter,J.C.
TITLE Use of BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.
  1. .103
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="RPCI-11-4A12"
  /clone_lib="RPCI-11"
  /sex="Male"
  /cell_type="Lymphocytes"
  /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC library"
  30 a 28 c 30 g 15 t
BASE COUNT
ORIGIN
```

```

QY 12724 TTTTATAGACAGCGGGTTTACCATTGTTAGCCAGGATGGTCTCGATCTCTGACCTCG 12783
|||||
Db 99 TTTTATAGACAGCGGGTTTACCCTTTAGCCGGATGGTCTCGATCTCTGACCTCG 40
|||||

QY 12784 TGAATCGCCCACTGAGCCTCCCAAGTCTCGGATTAC 12822
|||||
Db 39 TGAATCGCCCGCTCGCGCTCCCAAGTCTCGGCTTAC 1
|||||

RESULT 12
AA565533 107 bp mRNA EST 08-SEP-1997
LOCUS n42b11.s1 NCI_CGAP_GC2 Homo sapiens cDNA clone IMAGE:1016157 3'
DEFINITION similar to contains Alu repetitive element,, mRNA sequence.
ACCESSION AA565533
VERSION AA565533.1 GI:2337172
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 107)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1393355.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Stratagene, Inc., David B. Krizman,
Ph.D.
DNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1661 Std Error: 0.00
Seq primer: -40m13 fwd. Et from Amersham
High quality sequence stop: 87.
FEATURES
Location/Qualifiers
1..107
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1016157"
/clone_lib="NCI_CGAP_GC2"
/tissue_type="germ cell tumor"
/lab_host="SOLR (kanamycin resistant)"
/Note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dt. Bulk
germ cell tumor. 5' adaptor sequence: 5' GAATTCGGCAGCAG 3'
3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'
Average insert size: 1.2 kb."
BASE COUNT 22 a 34 c 26 g 25 t
ORIGIN

Query Match 0.4%; Score 87.8; DB 35; Length 107;
Best Local Similarity 88.8%; Pred. No. 1.1;
Matches 95; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 14105 CTGGAGTCAGTGGCGGATCTCGGCTCACTGCAACCTCCGCTCCCGGGTTCAAGTGAT 14164
|||||
Db 1 CTGGAGTCAGTGGCTCACTCAGCTCACTGCAAGCTCTGCTCCAGGTTCAAGTGAT 60
|||||

QY 14165 TCTCCTGCCTCAGACTCCCGAGTAGCTGGGATTACAGTCATGCACCA 14211
|||||
Db 61 TCTCCTGCCTCAGCTCTCTGAGTAGTGGGATTACAGGCACACCA 107
|||||

```

```

RESULT 13
AA228795 103 bp mRNA EST 20-AUG-1997
LOCUS nc14e07.s1 NCI_CGAP_Pr1 Homo sapiens cDNA clone IMAGE:1008132
DEFINITION similar to contains Alu repetitive element; contains element MER28
repetitive element ;, mRNA sequence.
ACCESSION AA228795
VERSION AA228795.1 GI:1851455
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 103)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1394473.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuauqui,
M.D., Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
DNA Library Arraying by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -41m13 fwd. Et from Amersham
High quality sequence stop: 81.
FEATURES
Location/Qualifiers
1..103
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1008132"
/clone_lib="NCI_CGAP_Pr1"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/Note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected, histologically normal
prostate epithelial cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into pAMP10 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Krizman."
BASE COUNT 14 a 30 c 25 g 34 t
ORIGIN

Query Match 0.4%; Score 87; DB 30; Length 103;
Best Local Similarity 90.3%; Pred. No. 1.4;
Matches 93; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 3786 TTTTATTTTTCAGACGGTGTCTACTCTTCGCCCGCCGCGGACTGCAGTAGCCGTAT 3845
|||||
Db 1 TTTTATTTTTCAGATGTGTCTACTCTGTGCGCCGAGGCTGGAGTGCACTAGCAAT 60
|||||

QY 3846 CTCGGTCTACTGCAAGCTCCGCTCCCGGTTTCAAGCCATTTT 3888
|||||
Db 61 CTGGGCTACTGCAAGCTCCGCTCCCGGTTTCAAGCCGTAT 103
|||||

RESULT 14
AQ582186

```



---







Matches 87; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 12715 TTTTGTATTTTGTAGACAGCGGGTTTCACCATGTTAGCCAGGATGGTCTCGATCTC 12774

Db 1 TTTTGTATCTTTAGTAGAGACAGGGTTTCACCATATGTCAGGCTGCTCTCAAACTC 60

Qy 12775 CTGACCTCGTGATCGGCCACCTGAGCCTCCCAAAGTCTGGGAT 12819

Db 61 CTGACCTGTGTATCACCAGCCTCGGCCCTCCCAAAGTCTGGGAT 105

## RESULT 2

US-08-477-504A-65

; Sequence 65, Application US/08477504A

; Patent No. 5972353

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

; CITY: Tiburon

; STATE: California

; COUNTRY: USA

; ZIP: 94920

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/477,504A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/260,190

; FILING DATE: 15-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lauder, Leona L.

; REGISTRATION NUMBER: 30,863

; REFERENCE/DOCKET NUMBER: D-0021.3D

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-435-2034

; TELEFAX: 415-435-0727

; INFORMATION FOR SEQ ID NO: 65:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 105 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-477-504A-65

Query Match 0.3%; Score 76.2; DB 4; Length 105;  
Best Local Similarity 82.9%; Pred. No. 4.9e-06;  
Matches 87; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 12715 TTTTGTATTTTGTAGACAGCGGGTTTCACCATGTTAGCCAGGATGGTCTCGATCTC 12774

Db 1 TTTTGTATCTTTAGTAGAGACAGGGTTTCACCATATGTCAGGCTGCTCTCAAACTC 60

Qy 12775 CTGACCTCGTGATCGGCCACCTGAGCCTCCCAAAGTCTGGGAT 12819

Db 61 CTGACCTGTGTATCACCAGCCTCGGCCCTCCCAAAGTCTGGGAT 105

## RESULT 3

US-08-486-756A-65

; Sequence 65, Application US/08486756A

; Patent No. 5981711

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

; CITY: Tiburon

; STATE: California

; COUNTRY: USA

; ZIP: 94920

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/486,756A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/260,190

; FILING DATE: 15-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lauder, Leona L.

; REGISTRATION NUMBER: 30,863

; REFERENCE/DOCKET NUMBER: D-0021.3C

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-435-2034

; TELEFAX: 415-435-0727

; INFORMATION FOR SEQ ID NO: 65:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 105 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-486-756A-65

Query Match 0.3%; Score 76.2; DB 4; Length 105;  
Best Local Similarity 82.9%; Pred. No. 4.9e-06;  
Matches 87; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 12715 TTTTGTATTTTGTAGACAGCGGGTTTCACCATGTTAGCCAGGATGGTCTCGATCTC 12774

Db 1 TTTTGTATCTTTAGTAGAGACAGGGTTTCACCATATGTCAGGCTGCTCTCAAACTC 60

Qy 12775 CTGACCTCGTGATCGGCCACCTGAGCCTCCCAAAGTCTGGGAT 12819

Db 61 CTGACCTGTGTATCACCAGCCTCGGCCCTCCCAAAGTCTGGGAT 105

## RESULT 4

US-08-485-862B-65

; Sequence 65, Application US/08485862B

; Patent No. 5989838

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

```

; APPLICATION NUMBER: US/08/787,739
; FILING DATE: 24-JAN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,049
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/486,756
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/477,504
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/481,658
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,862
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,863
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,077
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-981-2034
; TELEFAX: 415-981-0332
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-787-739-65

Query Match 0.3%; Score 76.2; DB 5; Length 105;
Best Local Similarity 82.9%; Pred. No. 4.9e-06;
Matches 87; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 12715 TTTTGTGATTTTGTAGTAGACACGGGGTTTCCACCATGTTAGCCAGGATGGTCTCGATCTC 12774
      |||||
Db 1 TTTTTCATCTTTTAGTAGACAGAGGGTTTCCACCATATTTGGCCAGGCTCTCAAATC 60
      |||||

QY 12775 CTGACCTGTGATCGCCACCTGAGCCCTCCAAAGTCTGGGAT 12819
      |||||
Db 61 CTGACCTGTGATCCACCGAGCTGGGCTCCCAAAGTCTGGGAT 105
      |||||

RESULT 6
US-08-454-557C-91
; Sequence 91 Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: Of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:

```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,557C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; US-08-454-557C-91

Query Match      0.3%; Score 67; DB 3; Length 84;
Best Local Similarity 88.0%; Pred. No. 0.00023;
Matches 73; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 12747 CCATGTTAGCCAGGATGCTCGATCTCCTGACCTCGTGATCCGCCACCTGAGGCTCCC 12806
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1 CCATGTTTCATCAGGCTGGTGTCGAACCTCCTGACCTCGTGATCCGCCACCTGAGGCTCCC 60

QY 12807 AAAGTGCTGGGATTACAGGTGTG 12829
      ||||| ||||| ||||| ||||| |||||
Db 61 AAAGTGCTGGGATTACAGCGGTG 83

RESULT 7
US-08-340-426D-91
; Sequence 91, Application US/08340426D
; Patent No. 5948634
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/340,426D
; FILING DATE: 14-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
```

```
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; US-08-340-426D-91

Query Match      0.3%; Score 67; DB 4; Length 84;
Best Local Similarity 88.0%; Pred. No. 0.00023;
Matches 73; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 12747 CCATGTTAGCCAGGATGCTCGATCTCCTGACCTCGTGATCCGCCACCTGAGGCTCCC 12806
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1 CCATGTTTCATCAGGCTGGTGTCGAACCTCCTGACCTCGTGATCCGCCACCTGAGGCTCCC 60

QY 12807 AAAGTGCTGGGATTACAGGTGTG 12829
      ||||| ||||| ||||| ||||| |||||
Db 61 AAAGTGCTGGGATTACAGCGGTG 83

RESULT 8
US-08-450-673C-91
; Sequence 91, Application US/08450673C
; Patent No. 5948888
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,673C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; US-08-450-673C-91

Query Match      0.3%; Score 67; DB 4; Length 84;
Best Local Similarity 88.0%; Pred. No. 0.00023;
Matches 73; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 12747 CCATGTTAGCCAGGATGCTCGATCTCCTGACCTCGTGATCCGCCACCTGAGGCTCCC 12806
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1 CCATGTTTCATCAGGCTGGTGTCGAACCTCCTGACCTCGTGATCCGCCACCTGAGGCTCCC 60

QY 12807 AAAGTGCTGGGATTACAGGTGTG 12829
      ||||| ||||| ||||| ||||| |||||
Db 61 AAAGTGCTGGGATTACAGCGGTG 83
```

```
RESULT 9
PCT-US95-17111A-91
; Sequence 91, Application PC/TUS9517111A
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and
; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/17111A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/340,426
; FILING DATE: 14-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 0609.3840002
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
PCT-US95-17111A-91

Query Match 0.3%; Score 67; DB 6; Length 84;
Best Local Similarity 88.0%; Pred. No. 0.00023;
Matches 73; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 12747 CCATGTTAGCCAGGATGGTCTCGATCTCTGACCTGCTGATCGCGCCAGCTGAGCTCCC 12806
Db 1 CCATGTTATCATCAGGCTGGTGTGCAACTCTGACCTGCTGATCGCGCCAGCTGAGCTCCC 60

QY 12807 AAAGTCTGGGATTACAGGTGTG 12829
Db 61 AAAGTCTGGGATTACAGCGTG 83

RESULT 10
US-08-454-557C-91/c
; Sequence 91, Application US/0845457C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.

; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/340,426D
; FILING DATE: 14-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 0609.3840002
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2540
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both

US-08-454-557C-91

Query Match 0.2%; Score 60.6; DB 3; Length 84;
Best Local Similarity 83.1%; Pred. No. 0.0036;
Matches 69; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 7134 CATGCTCTTAATTCACGACCTTTGGGAGCGGAGGCGGAGGATCATGAGTTCAGGAATG 7193
Db 83 CACGCTTGTAATCCACGACCTTTGGGAGGCTGAGGCGGCGGATCAGGAGTTCAGGAGTT 24

QY 7194 CAAGACCAGCGCTGACCAATATGG 7216
Db 23 CGACACCAGCGCTGATGACATGG 1

RESULT 11
US-08-340-426D-91/c
; Sequence 91, Application US/08340426D
; Patent No. 5948634
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/340,426D
; FILING DATE: 14-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 0609.3840002
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2540
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
```

INFORMATION FOR SEQ ID NO: 91:

SEQUENCE CHARACTERISTICS:  
LENGTH: 84 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
US-08-340-426D-91

Query Match 0.2%; Score 60.6; DB 4; Length 84;  
Best Local Similarity 83.1%; Pred. No. 0.0036;  
Matches 69; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 7134 CATGCTGTAAATTCAGCACCTTTGGAGCGCGAGCGCAGATCATGAGGTTCAGGAATG 7193

Db 83 CACGCTTGTAAATCCAGCACCTTTGGAGCGCTGAGCGCGGATCAGGATCAGGAGTT 24

Qy 7194 CAAGACCAGCCTGACCAATATGG 7216

Db 23 CGACACCAGCCTGATGAACATGG 1

RESULT 12

US-08-450-673C-91/c  
Sequence 91, Application US/08450673C  
Patent No. 5948888

GENERAL INFORMATION:  
APPLICANT: de la Monte, Suzanne

APPLICANT: Wands, Jack R.  
TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection

TITLE OF INVENTION: of Alzheimer's Disease  
NUMBER OF SEQUENCES: 121

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

STREET: 1100 New York Avenue, Suite 500

CITY: Washington  
STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005-3934

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/450.673C

FILING DATE: 30-MAY-1995  
CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:  
NAME: Ludwig, Steven R.

REGISTRATION NUMBER: 36,203  
REFERENCE/DOCKET NUMBER: 0609.3840004

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 91:

SEQUENCE CHARACTERISTICS:  
LENGTH: 84 base pairs

TYPE: nucleic acid  
STRANDEDNESS: both

TOPOLOGY: both  
US-08-450-673C-91

Query Match 0.2%; Score 60.6; DB 4; Length 84;  
Best Local Similarity 83.1%; Pred. No. 0.0036;  
Matches 69; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 7134 CATGCTGTAAATTCAGCACCTTTGGAGCGCGAGCGCAGATCATGAGGTTCAGGAATG 7193

Db 83 CACGCTTGTAAATCCAGCACCTTTGGAGCGCTGAGCGCGGATCAGGATCAGGAGTT 24

Qy 7194 CAAGACCAGCCTGACCAATATGG 7216

Db 23 CGACACCAGCCTGATGAACATGG 1

RESULT 13

PCT-US95-17111A-91/c

Sequence 91, Application PC/TUS9517111A  
GENERAL INFORMATION:

APPLICANT: de la Monte, Suzanne

APPLICANT: Wands, Jack R.

TITLE OF INVENTION: Neural Thread Protein Gene Expression and

TITLE OF INVENTION: Detection of Alzheimer's Disease

NUMBER OF SEQUENCES: 121

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

STREET: 1100 New York Avenue, Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/17111A

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/340,426

FILING DATE: 14-NOV-1994

ATTORNEY/AGENT INFORMATION:

NAME: Ludwig, Steven R.

REGISTRATION NUMBER: 36,203

REFERENCE/DOCKET NUMBER: 0609.3840002

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 91:

SEQUENCE CHARACTERISTICS:

LENGTH: 84 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: both  
PCT-US95-17111A-91

Query Match 0.2%; Score 60.6; DB 6; Length 84;

Best Local Similarity 83.1%; Pred. No. 0.0036;

Matches 69; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 7134 CATGCTGTAAATTCAGCACCTTTGGAGCGCGAGCGCAGATCATGAGGTTCAGGAATG 7193

Db 83 CACGCTTGTAAATCCAGCACCTTTGGAGCGCTGAGCGCGGATCAGGATCAGGAGTT 24

Qy 7194 CAAGACCAGCCTGACCAATATGG 7216

Db 23 CGACACCAGCCTGATGAACATGG 1

RESULT 14

US-08-454-557C-70

Sequence 70, Application US/08454557C

Patent No. 5830670

GENERAL INFORMATION:

APPLICANT: de la Monte, Suzanne

APPLICANT: Wands, Jack R.

TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection

TITLE OF INVENTION: of Alzheimer's Disease

NUMBER OF SEQUENCES: 121

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

STREET: 1100 New York Avenue, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/454,557C  
FILING DATE: 30-MAY-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Ludwig, Steven R.  
REGISTRATION NUMBER: 36,203  
REFERENCE/DOCKET NUMBER: 0609.3840003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 70:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
US-08-454-557C-70

	Query Match	0.2%	Score 58.8;	DB 3;	Length 78;
	Best Local Similarity	84.6%;	Pred. No. 0.0075;		
	Matches 66;	Conservative 0;	Mismatches 12;	Indels 0;	Gaps 0;
QY 16969	ACACGCCCTGCTAAATTTTGTATTTCTGGTAGACGGGGTTTCACCATGTTGGCCAGG	17028			
Db 1	ACAACGCCCCAGCTAAATATTTCTATTTTCTAGTAGAGATGGGGCTTCTCCATGTTTCATCAGG	60			
QY 17029	CTGGTCTCGAACTCCTGA	17046			
Db 61	CTGGTGTCTCGAACTCCTGA	78			

Search completed: June 21, 2000, 05:22:58  
Job time: 577708 sec

RESULT 15  
US-08-340-426D-70  
; Sequence 70, Application US/08340426D  
; Patent No. 5948634  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; APPLICANT: Wands, Jack R.  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; TITLE OF INVENTION: of Alzheimer's Disease  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/340,426D  
; FILING DATE: 14-NOV-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, Steven R.  
; REGISTRATION NUMBER: 36,203  
; REFERENCE/DOCKET NUMBER: 0609.3840002





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2000, 12:38:36 ; Search time 17971.8 seconds  
(without alignments)  
-1569.738 Million cell updates/sec

Title: US-08-852-495C-1\_COPY\_1\_29000  
Perfect score: 29000  
Sequence: 1 CACACACACACACACACA.....CCAGGCTAGAGTCAGTGGC 29000

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 370290

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_bal.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pl1.\*
- 8: gb\_pl2.\*
- 9: gb\_pr1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_vi.\*
- 17: em\_fun.\*
- 18: em\_hum1.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_om.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pl.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_vi.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_bal.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*
- 44: gb\_htg6.\*

- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*
- 50: gb\_pl3.\*
- 51: gb\_pr5.\*
- 52: gb\_htg8.\*
- 53: gb\_htg9.\*
- 54: gb\_htg10.\*
- 55: gb\_htg11.\*
- 56: gb\_htg12.\*
- 57: gb\_htg13.\*
- 58: gb\_htg14.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
C 1	87.2	0.3	103	9	HUMALCE221	M87896 Human carci
C 2	87	0.3	107	9	HUMALCE162	M87924 Human carci
C 3	87.2	0.3	108	10	HSLDLRN2	X05250 Human LDL-r
C 4	83.6	0.3	108	10	HSLDLRN2	X05250 Human LDL-r
C 5	79.8	0.3	108	10	HSLDLRD1	X05249 Human LDL-r
C 6	79.8	0.3	108	10	HSLDLRD2	X05251 Human LDL-r
C 7	80	0.3	108	11	HSU67803	U67803 Human small
C 8	75	0.3	103	9	HUMALCE221	M87896 Human carci
C 9	75.2	0.3	108	11	HSU67804	U67804 Human small
C 10	74.2	0.3	108	9	HUMDL003M5	D16965 Human HepG2
C 11	73.6	0.3	108	10	HSLDLI12	X05248 Human LDL-r
C 12	73.4	0.3	110	11	HSU67807	U67807 Human small
C 13	73	0.3	103	13	HS8IC8R	X57789 Human sequ
C 14	73	0.3	108	10	HSLDLRD1	X05249 Human LDL-r
C 15	73	0.3	108	10	HSLDLRD2	X05251 Human LDL-r
C 16	72.4	0.2	101	10	S79560	S79560 HRX (intron
C 17	71.6	0.2	94	9	HUMHGAL	M13479 Human alpha
C 18	70.8	0.2	90	9	HUMDLRFL	K03555 Human low d
C 19	70.8	0.2	91	13	HUMUT8164A	L30244 Human STS U
C 20	69.8	0.2	108	13	G32614	G32614 A009K21 Hum
C 21	69.8	0.2	110	9	HUMALCE43	M87900 Human carci
C 22	68.8	0.2	106	13	G32743	G32743 A009P31 Hum
C 23	69	0.2	108	11	HSU67803	U67803 Human small
C 24	68.8	0.2	108	11	HSU67808	U67808 Human small
C 25	68.4	0.2	95	13	HUMUT8002B	L30176 Human STS U
C 26	68	0.2	97	9	HUMDLRLA1	M14178 Human low d
C 27	67.8	0.2	100	9	HUMGALNSA	D45223 Human GALNS
C 28	68	0.2	103	13	HS8IC8R	X57789 Human sequ
C 29	68	0.2	104	9	HUMALCE272	M87899 Human carci
C 30	68	0.2	107	9	HUMALCE162	M87924 Human carci
C 31	67.8	0.2	108	13	G43535	G43535 WIAF-2393-S
C 32	67.2	0.2	85	10	HUMHIS1PR	M26162 Homo sapien
C 33	67.2	0.2	97	9	HUMDLRA2	M14180 Human low d
C 34	66.4	0.2	97	9	HUMDLRA2	M14180 Human low d
C 35	65.8	0.2	99	13	HUMUT7692A	L30306 Human STS U
C 36	66	0.2	100	13	G43536	G43536 WIAF-2394-S
C 37	66	0.2	100	13	G43538	G43538 WIAF-2396-S
C 38	66	0.2	110	9	HUMALCE43	M87900 Human carci
C 39	65.6	0.2	80	9	HUMBERFAE	M36135 Human alpha
C 40	65.6	0.2	107	11	HSU67806	U67806 Human small
C 41	65.2	0.2	79	10	S73203	S73203 ALL-1 (tand
C 42	65.4	0.2	110	11	HSU67807	U67807 Human small
C 43	64.6	0.2	95	10	HSSTHPKIB	X66361 H.sapiens m
C 44	64.8	0.2	96	4	NVIHIS2A	J00950 Newt histon
C 45	64.4	0.2	97	9	HUMDLRDJ	M14179 Human faml1

ALIGNMENTS

```

RESULT 1
HUMALCE221/c
LOCUS HUMALCE221 103 bp ss-RNA PRI 15-APR-1994
DEFINITION Human carcinoma cell-derived Alu RNA transcript, clone CE221.
ACCESSION M87896
VERSION M87896.1 GI:174874
KEYWORDS Alu repeat.
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 103)
AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
post-transcriptional selection of master sequences
JOURNAL J. Mol. Biol. (1992) In press
FEATURES
    source
        1..103
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /cell_line="Ntera2D1"
            /dev_stage="embryo"
            /sex="male"
            /tissue_type="carcinoma"
BASE COUNT 25 a 27 c 33 g 18 t
ORIGIN
Query Match 0.3%; Score 87.2; DB 9; Length 103;
Best Local Similarity 92.0%; Pred. No. 9.1e-06;
Matches 92; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 12205 CTGGAGTCAATGGCGGCGATCTTGCTCACACACCTCCGCTCCGGGTTCAAGCCAT 12264
|||||
Db 103 CTGGAGTCAATGGCGGCGATCTTGCTCACACCTCCGCTCCGGGTTCAAGCCAT 44
|||||

QY 12265 TCTCTGCTCAGCTCCGAGTAGCTGGGATTACAGGCA 12304
|||||
Db 43 TCTCTGCTCAGCTCCGCTAGCTGGGATTACAGGCA 4
|||||

RESULT 2
HUMALCE162/c
LOCUS HUMALCE162 107 bp ss-RNA PRI 15-APR-1994
DEFINITION Human carcinoma cell-derived Alu RNA transcript, clone CE162.
ACCESSION M87924
VERSION M87924.1 GI:174871
KEYWORDS Alu repeat.
SOURCE Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 107)
AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
post-transcriptional selection of master sequences
JOURNAL J. Mol. Biol. (1992) In press
FEATURES
    source
        1..107
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /cell_line="Ntera2D1"
            /dev_stage="embryo"
            /sex="male"
            /tissue_type="carcinoma"
BASE COUNT 28 a 30 c 35 g 14 t
ORIGIN
Query Match 0.3%; Score 87; DB 9; Length 107;
Best Local Similarity 90.3%; Pred. No. 9.8e-06;
Matches 93; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```

```

QY 3564 TTTTGTGAGAGGAGCTAGCTCTGTCGCCAGGCTGGAGTGCAGTGGCACCATCTTGGC 3623
|||||
Db 107 TTTTGTGAGAGGAGTCTGCTCTGTCGCCAGGCTGGAGTGCAGTGGCAGCATCTCGGC 48
|||||

QY 3624 TCATGCAAGCTCTGCCCTCCCGGGTTATGCAATCTCATGTC 3666
|||||
Db 47 TCATGCAAGCTCCGCTCCCGGGTTACAGGCATCTTCTTGGC 5
|||||

RESULT 3
HSLDLRN2/c
LOCUS HSLDLRN2 108 bp DNA PRI 20-MAY-1992
DEFINITION Human LDL-receptor gene intron 14 fragment (normal gene).
ACCESSION X05250
VERSION X05250.1 GI:34337
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
Williamson,R. and Humphries,S.
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT See X05252 for deletion junction
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES
    source
        1..108
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            1..108
            /note="intron XIV fragment"
BASE COUNT 28 a 23 c 39 g 18 t
ORIGIN
Query Match 0.3%; Score 87.2; DB 10; Length 108;
Best Local Similarity 88.0%; Pred. No. 9.1e-06;
Matches 95; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 3618 CTGTGCTCACTGCACAGCTCTGCTCCCGGGTTATGCCATCTCATGCTCAGCCTCCAG 3677
|||||
Db 108 CTCGGCTCACTGCACACCTCTGCTCCCTGGGTTCAAGCAATCTCTGCTCAGCCTCCGG 49
|||||

QY 3678 AGTACTGGGACTACAGCGCCGCCACACACCGCTGGCTAATTTTTT 3725
|||||
Db 48 AGTACTGGGATTACAGGCACCTGCCACCACCGCTGGCTAATTTTTGT 1
|||||

RESULT 4
HSLDLRN2
LOCUS HSLDLRN2 108 bp DNA PRI 20-MAY-1992
DEFINITION Human LDL-receptor gene intron 14 fragment (normal gene).
ACCESSION X05250
VERSION X05250.1 GI:34337
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
Williamson,R. and Humphries,S.
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901

```

	Query Match	0.3%	Score 79.8;	DB 10;	Length 108;
	Best Local Similarity	84.1%;	Pred. No. 0.00018;		
	Matches 90;	Conservative	Mismatches 17;	Indels	Gaps 0;
Qy	3619	TTGGCTCACTCGAAGCTCTGCCTCCGCGTTTTATGCCATTCTCATGTCCTCAGCCTCCAGA	3678		
Db	2	TGCGCTCACCAACCTCTGCCTCTCTGGTTCAAACCATTTTCTGCCTCAGCCTCCCGA	61		
Qy	3679	GTAGCTGGGACTACAGGCGCCGCCACACGCCCTGGCTAATTTTTTTT	3725		
Db	62	GTAGCTGGGATTACAGGCACCCTGGCCACACGCCCTGGGTAATTTTGT	108		

REFERENCE AUTHORS TITLE	JOURNAL MEDLINE REFERENCE AUTHORS TITLE JOURNAL
1. K. A. B. ... J. ... ...	1. K. A. B. ... J. ... ...
2. ... ...	2. ... ...
3. ... ...	3. ... ...
4. ... ...	4. ... ...
5. ... ...	5. ... ...
6. ... ...	6. ... ...
7. ... ...	7. ... ...
8. ... ...	8. ... ...
9. ... ...	9. ... ...
10. ... ...	10. ... ...
11. ... ...	11. ... ...
12. ... ...	12. ... ...
13. ... ...	13. ... ...
14. ... ...	14. ... ...
15. ... ...	15. ... ...
16. ... ...	16. ... ...
17. ... ...	17. ... ...
18. ... ...	18. ... ...
19. ... ...	19. ... ...
20. ... ...	20. ... ...
21. ... ...	21. ... ...
22. ... ...	22. ... ...
23. ... ...	23. ... ...
24. ... ...	24. ... ...
25. ... ...	25. ... ...
26. ... ...	26. ... ...
27. ... ...	27. ... ...
28. ... ...	28. ... ...
29. ... ...	29. ... ...
30. ... ...	30. ... ...
31. ... ...	31. ... ...
32. ... ...	32. ... ...
33. ... ...	33. ... ...
34. ... ...	34. ... ...
35. ... ...	35. ... ...
36. ... ...	36. ... ...
37. ... ...	37. ... ...
38. ... ...	38. ... ...
39. ... ...	39. ... ...
40. ... ...	40. ... ...
41. ... ...	41. ... ...
42. ... ...	42. ... ...
43. ... ...	43. ... ...
44. ... ...	44. ... ...
45. ... ...	45. ... ...
46. ... ...	46. ... ...
47. ... ...	47. ... ...
48. ... ...	48. ... ...
49. ... ...	49. ... ...
50. ... ...	50. ... ...
51. ... ...	51. ... ...
52. ... ...	52. ... ...
53. ... ...	53. ... ...
54. ... ...	54. ... ...
55. ... ...	55. ... ...
56. ... ...	56. ... ...
57. ... ...	57. ... ...
58. ... ...	58. ... ...
59. ... ...	59. ... ...
60. ... ...	60. ... ...
61. ... ...	61. ... ...
62. ... ...	62. ... ...
63. ... ...	63. ... ...
64. ... ...	64. ... ...
65. ... ...	65. ... ...
66. ... ...	66. ... ...
67. ... ...	67. ... ...
68. ... ...	68. ... ...
69. ... ...	69. ... ...
70. ... ...	70. ... ...
71. ... ...	71. ... ...
72. ... ...	72. ... ...
73. ... ...	73. ... ...
74. ... ...	74. ... ...
75. ... ...	75. ... ...
76. ... ...	76. ... ...
77. ... ...	77. ... ...
78. ... ...	78. ... ...
79. ... ...	79. ... ...
80. ... ...	80. ... ...
81. ... ...	81. ... ...
82. ... ...	82. ... ...
83. ... ...	83. ... ...
84. ... ...	84. ... ...
85. ... ...	85. ... ...
86. ... ...	86. ... ...
87. ... ...	87. ... ...
88. ... ...	88. ... ...
89. ... ...	89. ... ...
90. ... ...	90. ... ...
91. ... ...	91. ... ...
92. ... ...	92. ... ...
93. ... ...	93. ... ...
94. ... ...	94. ... ...
95. ... ...	95. ... ...
96. ... ...	96. ... ...
97. ... ...	97. ... ...
98. ... ...	98. ... ...
99. ... ...	99. ... ...
100. ... ...	100. ... ...

eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia  
 1 (bases 1 to 108)  
 Shaikh, T.H., Roy, A.M., Kim, J., Batzer, M.A. and Deininger, P.L. L.  
 CDNAs derived from primary and small cytoplasmic Alu (scAlu)  
 transcripts  
 J. Mol. Biol. 271 (2), 222-234 (1997)  
 97415756  
 2 (bases 1 to 108)  
 Shaikh, T.H., Kim, J., Batzer, M.A. and Deininger, P.L.  
 Direct Submission  
 Submitted (22-AUG-1996) Human Genetics and Molecular Biology

Children's Hospital of Philadelphia, 1004F Abramson Research  
Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA

## FEATURES

source

1..108

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="TscAlu2"

1..108

/note="scAlu"

/rpt\_family="Alu"

/rpt\_type="dispersed"

repeat\_region 23 a 39 c 30 g 16 t

## BASE COUNT

ORIGIN

Query Match 0.3%; Score 80; DB 11; Length 108;

Best Local Similarity 89.6%; Pred. No. 0.00016;

Matches 86; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 3735 GTAGAGATGGGTTTCCACCGTGTAGCCAGAAATGCTCGATCTCTTGACCTTCGTGATCC 3794

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 97 GTAGAGACGGGTTTCCACCTGTTAGCCAGGATGCTCGATCTCTGACCTCGTGATCC 38

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 3795 GCCTGCTTGGCTTCCAAAGTCTCCAAAGTCTGGGATTACAG 3830

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 37 GCCCGCTCGGCTCCCAAGTCTGGGATTACAG 2

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

## RESULT 8

HUMALCE221

LOCUS

DEFINITION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..103

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/cell\_line="Ntera2D1"

/dev\_stage="embryo"

/sex="male"

/tissue\_type="carcinoma"

BASE COUNT 25 a 27 c 33 g 18 t

ORIGIN

Query Match

Best Local Similarity

Matches 84; Conservative

0.3%; Score 75; DB 9;

Pred. No. 0.0012;

Mismatches 15; Indels

0; Gaps 0;

QY 24547 GCCTATATCCAGCTAATTTGGAGGCTGAGCGAGGAGTAATTCCTGAACTGGGAGGTG 24606

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 5 GCCTGTATATCCAGCTACACGGGAAGTAAGCGAGGAGTAATTCCTGAACTGGGAGGTG 64

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 24607 GAGGTGCTACTGAGCAGATCACACCATTCGACATCCAG 24645

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 65 GAGGTGCTACTGAGCAGATCGTGCCATTCGACATCCAG 103

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 9

HSU67804/c

LOCUS

DEFINITION

Human small cytoplasmic Alu transcript.

108 bp RNA

PRI

01-AUG-1997

ACCESSION U67804

VERSION U67804.1

KEYWORDS GI:2289918

SOURCE Alu.

ORGANISM human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 108)

AUTHORS Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.

TITLE cDNAs derived from primary and small cytoplasmic Alu (scAlu)

transcripts

J. Mol. Biol. 271 (2), 222-234 (1997)

MEDLINE 97415756

REFERENCE 2 (bases 1 to 108)

AUTHORS Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.

TITLE Direct Submission

JOURNAL Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The

Children's Hospital of Philadelphia, 1004F Abramson Research

Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA

FEATURES Location/Qualifiers

source

1..108

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="TscAlu3"

1..108

/note="scAlu"

/rpt\_family="Alu"

/rpt\_type="dispersed"

repeat\_region 26 a 38 c 26 g 18 t

BASE COUNT

ORIGIN

Query Match 0.3%; Score 75.2; DB 11; Length 108;

Best Local Similarity 86.5%; Pred. No. 0.0011;

Matches 83; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 3735 GTAGAGATGGGTTTCCACCGTGTAGCCAGAAATGCTCGATCTCTTGACCTTCGTGATCC 3794

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 97 GGAAGACGGGTTTCCACCTGTTAGCCAGGATGCTCGATCTCTGACCTTCGTGATCC 38

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 3795 GCCTGCTTGGCTTCCAAAGTCTCCAAAGTCTGGGATTACAG 3830

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 37 TCCCGCTTGGCTTCCAAAGTCTGGGATTACAG 2

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 10

HUMD1D03M5/c

LOCUS HUMD1D03M5 108 bp mRNA

DEFINITION Human HepG2 partial cDNA, clone hmd1d03m5.

ACCESSION D16965

VERSION D16965.1

KEYWORDS GI:598552

SOURCE gene signature.

ORGANISM Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 108)

AUTHORS Matoba,R.

TITLE Direct Submission

JOURNAL Submitted (21-JUL-1993) to the DBJ/EMBL/GenBank databases. Ryo

Matoba, Osaka University, Institute for Molecular and Cellular Bio;

1-3, Yamada-oka, Suita, Osaka 565, Japan

(E-mail:matoba@inherit.imcb.osaka-u.ac.jp,

Tel:81-6-877-5111(ex.3314), Fax:81-6-877-1922)

REFERENCE 2 (bases 1 to 108)

AUTHORS Matoba,R., Okubo,K., Hori,N., Fukushima,A. and Matsubara,K.

TITLE The addition of 5'-coding information to a 3'-directed cDNA library

improves analysis of gene expression

JOURNAL Gene 146 (2), 199-207 (1994)

MEDLINE 94357437

COMMENT Submitted (21-Jul-1993) to DDBJ by:

Ryo Matoba

Molecular Microbiology and Genetics Lab.

Research Institute of Innovative Technology for the Earth 9-2  
Kizugawadai Kizu-cyo,  
Soraku-gun, Kyoto  
Japan, 619-02  
Phone: 07747-5-2308  
Fax: 07747-5-2321.

## FEATURES

source  
1. .108  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Hep62"  
/clone\_lib="Kiseru"  
/sex="Male"

BASE COUNT 28 a 23 c 38 g 17 t 2 others  
ORIGIN

Query Match 0.3%; Score 74.2; DB 9; Length 108;  
Best Local Similarity 85.3%; Pred. No. 0.0017;  
Matches 93; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 12222 GATCTGGCTCAGCAACCTCCGCTCCCGGGTTCAAGCCATTCCTGCTCAGCCCTC 12281  
|||||  
Db 108 GATCTGGCTCAGCAACCTCCGCTCCCGGGTTCAAGCCATTCCTGCTCAGCCCTC 49  
|||||

QY 12282 CGGAGTAGCTGGGATTACAGGCAATCGCCACGACACCTCGCTAAATTTT 12330  
|||||

Db 48 CTGAGTAGCTGGGATTACACGATGGCCACACACCTGCTGCTTTTAT 1  
|||||

## RESULT 11

HSDDL112 108 bp DNA PRI 20-MAY-1992  
LOCUS Human LDL-receptor gene intron 12 fragment (normal gene) LDL = low density lipoprotein.

ACCESSION X05248  
VERSION X05248.1 GI:34334  
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor; repetitive sequence.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 108)  
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,

TITLE Unequal crossing-over between two alu-repetitive DNA sequences in the low-density-lipoprotein-receptor gene. A possible mechanism for the defect in a patient with familial hypercholesterolaemia  
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)  
MEDLINE 87161901

COMMENT see X05249 for deletion junction  
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

## FEATURES

source  
1. .108  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
misc\_feature complement(1..65)  
/note="Alu repeat"  
intron 1. .108  
/note="intron XII fragment"

BASE COUNT 21 a 38 c 20 g 29 t  
ORIGIN

Query Match 0.3%; Score 73.6; DB 10; Length 108;  
Best Local Similarity 81.7%; Pred. No. 0.0021;  
Matches 85; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 12226 TTGGCTCAGCAACCTCCGCTCCCGGGTTCAAGCATTTCTCTGCTCAGCCCTCGGA 12285  
|||||

Db 2 TCGCCTCACCACAACCTCTGCCCTCCCTGGGTTCAAGCATTTTCTGCTCAGCCCTCCTTA 61  
|||||

QY 12286 GTAGCTGGGATTACAGCATGCCACACACCTCGCTAAATTT 12329  
|||||

Db 62 GTAGCTGGGATTACAGCATGTGCCACACCGCCGCTGATTTT 105  
|||||

## RESULT 12

HSU67807/c 110 bp RNA PRI 01-AUG-1997  
LOCUS Human small cytoplasmic Alu transcript.

DEFINITION U67807  
ACCESSION U67807.1 GI:2289921  
VERSION U67807.1  
KEYWORDS Alu.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 110)  
AUTHORS Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.  
TITLE cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts

J. Mol. Biol. 271 (2), 222-234 (1997)  
MEDLINE 97415756

REFERENCE 2 (bases 1 to 110)  
AUTHORS Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.  
TITLE Direct Submission

JOURNAL Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The Children's Hospital of Philadelphia, 1004F Abramson Research Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA  
FEATURES  
1. .110  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="TscAlu6"  
repeat\_region 1. .110  
/note="scAlu"  
/rpt\_family="Alu"  
/rpt\_type="dispersed"

BASE COUNT 26 a 39 c 24 g 21 t  
ORIGIN

Query Match 0.3%; Score 73.4; DB 11; Length 110;  
Best Local Similarity 83.8%; Pred. No. 0.0023;  
Matches 83; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 2864 GTAGAGATGGGGTTTCACTATGTGGCCAGGCTAGTTTGAACCTCCTCAGCTGAT 2923  
|||||

Db 99 GGAAGATGGGGTTTCACTATGTGGCCAGGCTAGTTTGAACCTCCTCAGCTGAT 40  
|||||

QY 2924 CCATTCTCATTTGGCTCCCAAGTGTGGGATTACAGGC 2962  
|||||

Db 39 CCACCCACTTTGGCCCTCTCAAGTGTGGGATTACAGGC 1  
|||||

## RESULT 13

HS81C8R 103 bp DNA STS 05-SEP-1991  
LOCUS Human sequence tagged site 81C8R DNA from 19q13.

DEFINITION X57789  
ACCESSION X57789.1 GI:23938  
VERSION X57789.1  
KEYWORDS STS; myotonic dystrophy.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 103)  
AUTHORS Aldridge,F.L.

TITLE Direct Submission  
JOURNAL Submitted (12-FEB-1991) F.L. Aldridge, ICI Pharmaceuticals, Alderley Park, Macclesfield, Cheshire, SK10 4TG, UK

REFERENCE 2 (bases 1 to 103)  
AUTHORS Butler,R., Riley,J.H., Ogilvie,D.J., Anand,R., Buxton,J.,

Davies,J., Johnson,K. and Markham,A.F.













CC homologous non-human primate gene, in different cell types and/or at  
CC different developmental stages, using the non-primate transcript as  
CC probe is detected. A homologous gene having the same pattern of  
CC differential expression is selected and the non-primate gene, or part of  
CC it is used to identify the homologous human gene. The ESC transcripts  
CC identified by this method are used to generate transgenic animals  
CC baboon or chimpanzee for study of gene function. The method provides  
CC rapid and large scale screening for human developmental genes, and  
CC eliminates the need to analyse reporter gene expression in embryos.  
SQ Sequence 86 BP; 16 A; 28 C; 30 G; 12 T;

Query Match 0.2%; Score 65.4; DB 1; Length 86;  
Best Local Similarity 86.7%; Pred. No. 0.09;  
Matches 72; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 9626 TGACTCTCGTCTTCTTAGGAGACGCGCTGGATTTAGGAGGACGCGCCCTGA 9685

DB 84 TGGCTCTCCGCTCTCTTTGGGACGACGCGCTGGATTTGGCAGGACGCGCCCTGC 25

QY 9686 GCAATGGTCACCGCCTAGCAG 9708

DB 24 GCGATGGTCACGCGCCAGCAG 2

## RESULT 5

XI2087/c

ID XI2087 standard; DNA; 100 BP.

AC XI2087;

DE Human biallelic polymorphic DNA fragment EST98276a.

KW Polymorphism; biallelic; human; forensic; paternity testing; disease;

KW detection; phenotypic typing; characteristic; infection; hereditary;

KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;

KW treatment; marker; ss.

OS Homo sapiens.

PN W09820165-A2.

PD 14-MAY-1998.

PF 05-NOV-1997; U20313.

PR 06-NOV-1996; US-030455.

PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.

PI Hudson T, Lander ES, Wang D;

DR WPI; 98-286974/25.

PT New isolated nucleic acid segments from the human genome - used for

PT determining polymorphic forms for use in e.g. forensics, paternity

PT testing or phenotypic typing for disease

PS Claim 1; Page 219; 310pp; English.

CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic

CC markers which have been isolated using the primers represented in

CC X09121-X10268. The base occupying the polymorphic site is indicated by

CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in

CC methods for determining polymorphic forms in an individual for use in

CC e.g. forensics, paternity testing or for phenotypic typing for diseases

CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,

CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial

CC hypercholesterolemia, polycystic kidney disease, hereditary

CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary

CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos

CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,

CC autoimmune diseases, inflammation, cancer, diseases of the nervous

CC system, infection by pathogenic microorganisms, and characteristics such

CC as longevity, appearance (e.g. baldness, obesity), strength, speed,

CC endurance, fertility, and susceptibility or receptivity to particular

CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid

CC segments can also be used to produce medicaments for the treatment or

CC prophylaxis of such diseases.

SQ Sequence 100 BP; 21 A; 25 C; 22 G; 31 T;

## Query Match

Best Local Similarity 0.2%; Score 65; DB 1; Length 100;

Matches 77; Conservative 1; Mismatches 21; Indels 0; Gaps 0;

QY 23012 GTGGCTCATGCTTAATCCAGCACTTTGAGAGGCTGAAGAGGAGGATCGCTTGATC 23071

DB 99 GTGACTCACCTATATATCTTGCACCTTTAGGAGGCTTAGGAGGAGGATGTTTGAAC 40

QY 23072 CGGGAGTTCAAGAGCATCTCTGGGCAACACAGCGAGACCC 23110

DB 39 CAGGAGCTCAAGACCAKCTCTGGGAAACATAGCAAGACTC 1

## RESULT 6

XI2087

ID XI2087 standard; DNA; 100 BP.

AC XI2087;

DE Human biallelic polymorphic DNA fragment EST98276a.

KW Polymorphism; biallelic; human; forensic; paternity testing; disease;

KW detection; phenotypic typing; characteristic; infection; hereditary;

KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;

KW treatment; marker; ss.

OS Homo sapiens.

PN W09820165-A2.

PD 14-MAY-1998.

PF 05-NOV-1997; U20313.

PR 06-NOV-1996; US-030455.

PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.

PI Hudson T, Lander ES, Wang D;

DR WPI; 98-286974/25.

PT New isolated nucleic acid segments from the human genome - used for

PT determining polymorphic forms for use in e.g. forensics, paternity

PT testing or phenotypic typing for disease

PS Claim 1; Page 219; 310pp; English.

CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic

CC markers which have been isolated using the primers represented in

CC X09121-X10268. The base occupying the polymorphic site is indicated by

CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in

CC methods for determining polymorphic forms in an individual for use in

CC e.g. forensics, paternity testing or for phenotypic typing for diseases

CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,

CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial

CC hypercholesterolemia, polycystic kidney disease, hereditary

CC spherocytosis, von Willebrand's disease, tuberous sclerosis, Ehlers-Danlos

CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos

CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,

CC autoimmune diseases, inflammation, cancer, diseases of the nervous

CC system, infection by pathogenic microorganisms, and characteristics such

CC as longevity, appearance (e.g. baldness, obesity), strength, speed,

CC endurance, fertility, and susceptibility or receptivity to particular

CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid

CC segments can also be used to produce medicaments for the treatment or

CC prophylaxis of such diseases.

SQ Sequence 100 BP; 21 A; 25 C; 22 G; 31 T;

## Query Match

Best Local Similarity 0.2%; Score 63.4; DB 1; Length 100;

Matches 76; Conservative 1; Mismatches 22; Indels 0; Gaps 0;

QY 17709 GGGCTTACTATGTTCCCGAGGCTGCTCAAACTCTGGCTTAAGTATCTCTCTGCC 17768

DB 1 GAGCTTGTGTTTTTCCCGAGGTTGCTTGTAGCTCTGTTTCAAAACAATCTCTCTTCC 60

QY 17769 TCAGCTCCCAATTTCTGGGATTACTAGTGAGTCAC 17807

DB 61 TAAGCTCTCTAAAGTCCCGAGGATTATAGGTGTGAGTCAC 99

## RESULT 7

XI2085

ID XI2085 standard; DNA; 100 BP.

AC XI2085;

DE Human biallelic polymorphic DNA fragment EST98276c.

Polymorphism; biallelic; human; forensic; paternity testing; disease;  
detection; phenotypic typing; characteristic; infection; hereditary;  
autoimmune disease; cancer; inflammation; drug; therapy; medication;  
treatment; marker; ss.  
Homo sapiens.  
W09B20165-A2.  
PN W09B20165-A2.  
PD 14-MAY-1998.

Pf 05-NOV-1997; U20313.  
PF 06-NOV-1996; US-030455.  
PR (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
PA Hudson T, Lander ES, Wang D;  
PI WPI: 98-286974/25.  
DR New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease

CS Claim 1; Page 218; 310pp; English.  
PS X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
CC markers which have been isolated using the primers represented in  
CC X09121-X10268. The base occupying the polymorphic site is indicated by  
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberculous sclerosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.

CS Sequence 100 BP; 22 A; 25 C; 22 G; 30 T;

Query Match 0.2%; Score 63.4; DB 1; Length 100;  
Best Local Similarity 76.8%; Pred. No. 0.18;  
Matches 76; Conservative 1; Mismatches 22; Indels 0; Gaps 0

QY 17709 GGCTCTTACTATGTGGCCAGCGTGCTCAAACTCCGCGCTTAAGTCATGCCCTGC 17768  
I | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1 GAGCTCTGTATGTTTCCCAGGATGGCTTGAGCTCCTGCTTTCAAACAATCCTCCTTC 60

QY 17769 TCAGCCCTCCCAATGCTTGGGATTACTAGTGTGATCAC 17807  
I | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 61 TAAGCTCCCYAAGTCCAGGATTATAGTGTGATGTAC 99

RESULT 8

X12086

ID X12086 standard; DNA; 100 BP.

AC X12086;

DT 30-MAR-1999 (first entry)

DE Human biallelic polymorphic DNA fragment EST98276b.

DW Polymorphism; biallelic; human; forensic; paternity testing; disease;  
KW detection; phenotypic typing; characteristic; infection; hereditary;  
KW autoimmune disease; cancer; inflammation; drug; therapy; medication;  
KW treatment; marker; ss.  
KW Homo sapiens.  
PN W09B20165-A2.  
PD 14-MAY-1998.

Pf 05-NOV-1997; U20313.  
PF 06-NOV-1996; US-030455.  
PR (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
PA Hudson T, Lander ES, Wang D;  
PI WPI: 98-286974/25.  
DR New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease

CS Claim 1; Page 219; 310pp; English.

QY 9664 GTTAGGAGGACGC 9677  
DB 14 GTTGGCAGGACG 1

RESULT 10  
X12095/c

ID X12095 standard; DNA; 108 BP.  
AC X12095;  
DT 30-MAR-1999 (first entry)  
DE Human biallelic polymorphic DNA fragment TIGR-A003M18a.  
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;  
KW detection; phenotypic typing; characteristic; infection; hereditary;  
KW autoimmune disease; cancer; inflammation; drug; therapy; medication;  
KW treatment; marker; ss.  
OS Homo sapiens.  
PN W09820165-A2.  
PD 14-MAY-1998.  
PF 05-NOV-1997; U20313.  
PR 06-NOV-1996; US-030455.  
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
PI Hudson T, Lander ES, Wang D;  
DR WPI; 98-286974/25.  
PT New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease  
PS Claim 1; Page 219; 310pp; English.  
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
CC markers which have been isolated using the primers represented in  
CC X09121-X10268. The base occupying the polymorphic site is indicated by  
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, Ehlers-Danlos  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.  
SQ Sequence 108 BP; 19 A; 23 C; 28 G; 37 T;

Query Match 0.2%; Score 62.2; DB 1; Length 108;

Best Local Similarity 79.4%; Pred. No. 0.27; Mismatches 2; Gaps 1;  
Matches 85; Conservative 1; Mismatches 19; Indels 2; Gaps 1;

QY 24415 TGTATTTCCAGCACTTT--GGAGGCGAGCGCGGCGAGATCATTGAGTGGGAGTTCGA 24472  
DB 107 TATAATCCCGACACTTTTGGGAGGCGCAAGCGACGACGATCATTGAAGTCAGGAGTTCGA 48

QY 24473 GACTAGCTGGCCAAATGATGAAACCCCATCTCTACTAAAATACA 24519  
DB 47 GACCATCTCTGGCCAAACAYAGGAACCTCATCTCTACAAAAGACAA 1

RESULT 11  
T24892/c

ID T24892 standard; cDNA to mRNA; 100 BP.  
AC T24892;  
DT 05-NOV-1996 (first entry)  
DE Human gene signature HUMGS06998.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN W09514772-A1.

PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PI (OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI; 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1; Page 1720; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 100 BP; 28 A; 22 C; 25 G; 22 T;

Query Match 0.2%; Score 60.8; DB 1; Length 100;  
Best Local Similarity 74.7%; Pred. No. 0.43;  
Matches 74; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 3559 TTTTGTGAGCGGAGCTGTCTGTCGCCAGGCTGAGTGCACTGGCACCATC 3618  
DB 100 TTTGTTGTTTCAACAGAGTGTCACTGTCCACAGGCGAGTGCAANGTGCAATC 41

QY 3619 TTGGCTCACTGCAAGCTGTCCTCCCGGTTTATGCCAT 3657  
DB 40 TCAGCTNATTGCAAAATCTGCCTCCCGAGGTCAANGCGAT 2

RESULT 12  
T20743/c

ID T20743 standard; cDNA to mRNA; 102 BP.  
AC T20743;  
DT 26-JUL-1996 (first entry)  
DE Human gene signature HUMGS01961.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN W09514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PI (OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI; 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1; Page 714; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-



PA (MATS/) MATSUBARA K.  
PA (ORUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI; 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1: Page 1942; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 84 BP; 33 A; 17 C; 15 G; 19 T;

Query Match 0.2%; Score 56.4; DB 1; Length 84;  
Best Local Similarity 80.5%; Pred. No. 1.9;  
Matches 66; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 17679 TAATTTTAAAGCTTTTCTAGAGATGGGGTCTTACTATGTTGCCAGGCTGCTC 17738

Db 82 TAATTTTAAAGCTTTTCTAGAGATGGGGTCTTACTATGTTGCCAGGCTGCTC 23

QY 17739 AAACCTCCTGGGCTTAAGTGATC 17760

Db 22 GAACCTCCTGGGCTCAAGGATC 1

Search completed: June 14, 2000, 20:42:29  
Job time: 27860 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: June 14, 2000, 12:37:13 ; Search time 8513.17 Seconds  
(without alignments)  
13807.251 Million cell updates/sec

Title: US-08-852-495C-1\_COPY\_1\_29000  
Perfect score: 29000  
Sequence: 1 CACACACACACACACACACA.....CCAGGCTAGAGTCAGTGGC 29000

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 156056

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : EST:  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

Result Query  
SUMMARIES



clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 727 Std Error: 0.00

Seq primer: M13RP1

High quality sequence stop: 68.

Location/Qualifiers

## FEATURES

source

1. .95  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Soares fetal liver spleen INFUS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACGGGAAGTAATTAAGATCTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 31 a 23 c 26 g 12 t 3 others  
ORIGIN

Query Match 0.3%; Score 92; DB 21; Length 95;  
Best Local Similarity 96.8%; Pred. No. 0.066;  
Matches 92; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 10316 TGTGTTAAGATGCTCAGCAGCAGCAAGTCTGCTCCGCCGCGGAGAGGGCTCCAAAG 10375  
Db 1 TGTGTTAAGATGCTCAGCAGCAGCAAGTCTGCTCCGCCGCGGAGAGGGCTCCAAAG 60

QY 10376 GCAGTGACCAAGCGCAGAGAAGATGGCAAGAA 10410  
Db 61 GCAGTGACCAAGCGCAGAGAAGATGGCAAGAA 95

RESULT 3  
AI832832 105 bp mRNA EST 13-JUL-1999  
LOCUS at72g09.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone  
DEFINITION IMAGE:2377600.3, similar to contains Alu repetitive element; contains element MER22 repetitive element ;, mRNA sequence.  
ACCESSION AI832832  
VERSION AI832832.1 GI:5454812  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 105)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizan, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R. WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT On Dec 20, 1995 this sequence version replaced gi:1133644.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -400P from Glibco.

## FEATURES

source

Location/Qualifiers

1. .105

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Barstead colon HPLRB7"  
/sex="male"  
/dev\_stage="adult, age 25"  
/lab\_host="DH10B (phage resistant)"  
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTACGAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [5' AATCACTAGTAAT 3' and 5' ATTACTAGTG 3'], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead."

BASE COUNT 17 a 35 c 27 g 26 t  
ORIGIN

Query Match 0.3%; Score 92.2; DB 61; Length 105;  
Best Local Similarity 92.4%; Pred. No. 0.06;  
Matches 97; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 12176 GAGACCAAGTTCTCTCTGTTTCCAGGCTGGAGTGCATGGCGGATCTTGGCTCAC 12235  
Db 1 GAGACAGAGTTTGGCTCTGTTTCCAGGCTGGAGTGCATGGCGGATCTTGGCTCAC 60

QY 12236 GCACCTCCGCTCCGGGTTCAAGCAATTCCTCGCTCAGCCT 12280  
Db 61 GCACCTCCGCTCCGGGTTCAAGCAATTCCTCGCTCAGCCT 105

## RESULT 4

AA807640

LOCUS

DEFINITION

AA807640

ACCESSION

VERSION

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 103)

AUTHORS

TITLE

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

Unpublished (1997)

JOURNAL

COMMENT

On Jan 19, 1998 this sequence version replaced gi:2151346.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/dbp/image/image.html

Insert Length: 774 Std Error: 0.00

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 87.

Location/Qualifiers

1. .103

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="IMAGE:1255473"

/clone\_type="NCI-CGAP\_GC3"

/tissue\_type="pooled germ cell tumors"

```
/lab_host="DH10B"
/notice=vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled
germ cell tumors, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified p7T3
vector. Library is not normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      19 a      27 c      30 g      27 t
ORIGIN

Query Match      0.3%; Score 87.6; DB 38; Length 103;
Best Local Similarity 91.2%; Pred. No. 0.21;
Matches 93; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3734 AGTAGAGATGGGGTTTACCGGTGTTAGCAGATGGTCTCGATCTCTCGACCTTGTGATC 3793
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 AGTAGAGATGGGGTTTACCGGTGTTAGCAGATGGTCTCGATCTCTCGACCTTGTGATC 61

QY 3794 CGCCTGGCTTGCCTCCCAAGTCTGGGATACACGTGTGA 3835
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 CGCTCACCTCGGCTCCCAAGTCTGGGATACACGTGTGA 103

RESULT 5
T77382/c
LOCUS T77382 103 bp mRNA EST 15-MAR-1995
DEFINITION yd72h12.r1 Soares fetal liver spleen lNFLS Homo sapiens cDNA clone
IMAGE:113831 5' similar to contains Alu repetitive element; mRNA
sequence.
ACCESSION T77382
VERSION T77382.1 GI:694585
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 103)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Lultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Other_Ests: yd72h12.s1
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 943
Source: IMAGE Consortium, LLNL This clone is available royalty-free
through LLNL; contact the IMAGE Consortium (info@image.llnl.gov)
for further information. Putative full length read
Insert Length: 943 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 109.
Location/Qualifiers
1. 103
/organism="Homo sapiens"
/db_xref="db:469448"
/db_xref="taxon:9606"
/clone="IMAGE:113831"
/clone_lib="Soares fetal liver spleen lNFLS"
/sex="male"
/dev_stage="20 week post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/notice="Organ: Liver and Spleen; Vector: p7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
```

```
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACTGGAAGAATTAAATAAGATCTTTTTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      24 a      20 c      37 g      22 t
ORIGIN

Query Match      0.3%; Score 86.6; DB 21; Length 103;
Best Local Similarity 95.7%; Pred. No. 0.28;
Matches 89; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6077 TGAGAGTCTCACTCTCACTGCAACCTCCCTCTCTATATTTCAAGTGATCTCTTGCCTCA 6136
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 103 TGAGAGTCTCACTCTCACTGCAACCTCCCTCTCTATATTTCAAGTGATCTCTTGCCTCA 44

QY 6137 GCCTCCCGAGTAGCTGGGACTACAGCGCTGCAC 6169
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 43 GCCTCCCGAGTAGCTGGGACTACAGCGCTGCAC 11

RESULT 6
AA158786/c
LOCUS AA158786 106 bp mRNA EST 09-MAR-1998
DEFINITION z663cll.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
IMAGE:591572 5' similar to contains Alu repetitive element; contains
element PTR7 repetitive element; mRNA sequence.
ACCESSION AA158786
VERSION AA158786.1 GI:1733588
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 106)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
On Sep 12, 1996 this sequence version replaced gi:1406940.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Putative full length read
The vector to vector length is 119
Insert Length: 926 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham.
Location/Qualifiers
1. 106
/organism="Homo sapiens"
/db_xref="db:4622958"
/db_xref="taxon:9606"
/clone="IMAGE:591572"
/clone_lib="Stratagene pancreas (#937208)"
/lab_host="SOLR cells (kanamycin resistant)"
/notice="Organ: pancreas; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Pancreatic adenocarcinoma cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5'
CTCGAGCTTTTCTTTTTTTTTTTT 3'
BASE COUNT      27 a      28 c      37 g      14 t
ORIGIN
```

```
Db 49 TACTAAACTACAAAAATTAGCCGGCATGAAGGAGCATGACTGTAATC 1

RESULT 8
AA244245 110 bp mRNA EST 20-AUG-1997
LOCUS nc07a04.s1 NCI_CGAP_Prl Homo sapiens cDNA clone IMAGE:1007406
DEFINITION similar to contains Alu repetitive element,, mRNA sequence.
ACCESSION AA244245
VERSION AA244245.1 GI:1875104
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 110)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITILE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
On Jan 24, 1995 this sequence version replaced gi:634306.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
M.D., Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Cloned distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -41ml3 fwd. ET from Amersham
High quality sequence stop: 90.
Location/Qualifiers
1. .110
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1007406"
/clone_lib="NCI_CGAP_Prl"
/sex="Male"
/lab_host="DH10B"
/notes="Vector: PAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(GT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected, histologically normal
prostate epithelial cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into pAMP10 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Krizman."
BASE COUNT 17 a 26 c 28 g 38 t
ORIGIN

Query Match 0.3%; Score 85; DB 30; Length 110;
Best Local Similarity 85.5%; Pred. No. 0.42;
Matches 94; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 3560 TTTTCTTTTGTGACGAGCTAGCTCTGTGCGCCAGCTGGAGTCAGTGGCACCATCT 3619
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 TTTTCTTTTGTGAGATGGAGCTGTGATCTGTGCGCCAGCTGGAGTCAGTGGCAGANTCT 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 3620 TGGCTCAGTCAGAGCTCGCTCCCGGGTTATGCCATTCTCATGTCCTCA 3669
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 TGGCTCAGTCAGACCTCTGCCTCTCTGGGTTCAAGAGATTCTTCTGCTCTCA 110
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9

Query Match 0.3%; Score 84.8; DB 29; Length 106;
Best Local Similarity 88.5%; Pred. No. 0.45;
Matches 92; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 12190 CTCTTGTTCAGGCTGGAGTGCATGCGGGATCTTGGCTCACAGCAACCTCCGCCTC 12249
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 104 CTCTTGTTCAGGCTGGAGTGCATGCGGACCTTGGCTCACTGCACCTCCGCCTC 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 12250 CCGGGTTCAAGCATCTCTCGCTCAGCTCCGGAGTAGCTGG 12293
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 44 CCAAGTTTAAGCATCTCTGCTGCCCGCCTCTCTGAGTGGCTGG 1

RESULT 7
AQ029690/c
LOCUS RPC111-41F18-TV RPCI-11 Homo sapiens genomic clone RPCI-11-41F18,
DEFINITION genomic survey sequence.
ACCESSION AQ029690
VERSION AQ029690.1 GI:3274821
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 109)
AUTHORS Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K.,
Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and
Venter, J.C.
TITILE Use of BAC End Sequences for Sequence-Ready Map Building (1998)
JOURNAL Unpublished (1998)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamas@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1. .109
/organism="Homo sapiens"
/db_xref="GDB:751549"
/db_xref="taxon:9606"
/clone="RPCI-11-41F18"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/notes="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC Library"
BASE COUNT 21 a 26 c 24 g 38 t
ORIGIN

Query Match 0.3%; Score 85; DB 94; Length 109;
Best Local Similarity 86.2%; Pred. No. 0.42;
Matches 94; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 4703 CATATCACCTGAGGTGAGAGTTGAGACAGCTGGCCAAATGTGAACCCCTGCTC 4762
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 109 CAGATCACCTGAGGTGAGAGTTGAGACAGCTGGCCAAATGTGAACCCCTGATC 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 4763 TACTATAAATATAAAAAATTAGCTGGGTGGTGGTGCATGCTGTAGTC 4811
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

AA897366
LOCUS      AA897366      110 bp      mRNA      EST      04-JAN-1999
DEFINITION am06h02.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:1466067 3' similar to contains Alu repetitive element;; mRNA
sequence.

ACCESSION AA897366      1 GI:3033986
VERSION    AA897366
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 110)
AUTHORS    NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    On Jan 19, 1998 this sequence version replaced gi:2150764.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 834 Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 63.
FEATURES   Location/Qualifiers
            source          1..110
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:1466067"
                        /clone_11b="Soares_NFL_T_GBC_S1"
                        /lab_host="DH10B"
                        /note="Organ: pooled; Vector: pT7T3D-Pac (pharmacia) with
                        a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
                        Equal amounts of plasmid DNA from three normalized
                        libraries (fetal lung NBHL19W, testis NHT, and B-cell
                        NCI CGAP-GCBI) were mixed, and ss circles were made in
                        vitro. Following HAP purification, this DNA was used as
                        tracer in a subtractive hybridization reaction. The driver
                        was PCR-amplified cDNAs from pools of 5,000 clones made
                        from the same 3 libraries. The pools consisted of
                        I.M.A.G.E. clones 297480-302087, 682632-687239,
                        726408-728711, and 729096-731399. Subtraction by Bento
                        Soares and M. Fatima Bonaldo."
BASE COUNT 22 a 27 c 29 g 32 t
ORIGIN
Query Match 0.3%; Score 85; DB 39; Length 110;
Best Local Similarity 86.2%; Pred. No. 0.42;
Matches 94; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 3726 TTTATTATTAGTAGAGATGGGGTTTACCGGTGTACCGAGATGGTCTCGATCTCTTGACC 3785
+ + + + +
Db 2 TATTTTATTAGTAGATGGGGTTTACCGGTGTACCGAGATGGTCTCAATCTCTGGACC 61
+ + + + +

QY 3786 TTTCTATCGCGCTGCTTGGCTTCCCAAGTCTGGGATTACAGTGTG 3834
+ + + + +
Db 62 TCATGATCGCGCCACCTCGGCTCCCAAGTCTGGGATTATAGGCGTG 110
+ + + + +

RESULT 10
LOCUS      AO535244/c
DEFINITION RPCI-11-317H22.IV RPCI-11 Homo sapiens genomic clone
IMAGE:11-317H22, genomic survey sequence.
ACCESSION AO535244
VERSION    AO535244.1 GI:4846934
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 106)
AUTHORS    Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
            Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
            Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
            Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
            WashU-NCI human EST Project
            Unpublished (1997)
            On Sep 12, 1996 this sequence version replaced gi:1397630.
            Contact: Wilson RK
            Washington University School of Medicine

Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 103)
Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and
Venter, J.C.
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
Unpublished (1997)
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pleteredejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1..103
/organism="Homo sapiens"
/db_xref="GDB:7621533"
/db_xref="taxon:9606"
/clone="RPCI-11-317H22"
/clone_11b="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"
BASE COUNT 31 a 27 c 27 g 18 t
ORIGIN
Query Match 0.3%; Score 84.4; DB 108; Length 103;
Best Local Similarity 89.2%; Pred. No. 0.5;
Matches 91; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 12327 TTTTGTATTTTATAGACGACGAGTTTCTCCATGTCGTGAGTCTCGAACTCGC 12386
+ + + + +
Db 102 TTTTGTATTTTATAGACGACGAGTTTCTCCATGTTGCCAGGCTGTCGAACTCCT 43
+ + + + +

QY 12387 CACATCAGTGTATGTCGCCGCTTGGCTCCCAAGTCTCGG 12428
+ + + + +
Db 42 GACCTCAAGTGTATGTCGCCGCTTGGCTCCCAAGTCTCGG 1

RESULT 11
LOCUS      AA703692      106 bp      mRNA      EST      24-DEC-1997
DEFINITION ag81a10.r1 Stratagene hMT neuron (#937233) Homo sapiens cDNA clone
IMAGE:1140858 5' similar to contains Alu repetitive element;; mRNA
sequence.
ACCESSION AA703692
VERSION    AA703692.1 GI:2713610
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 106)
AUTHORS    Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
            Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
            Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
            Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
            WashU-NCI human EST Project
            Unpublished (1997)
            On Sep 12, 1996 this sequence version replaced gi:1397630.
            Contact: Wilson RK
            Washington University School of Medicine

```







Db 61 CATGAATTCCTTCGTCACGACATCNCATCAGCGCATC 97

Search completed: June 14, 2000, 15:22:36  
Job time: 9923 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: June 14, 2000, 12:54:30 ; Search time 372.68 seconds  
(without alignments)  
10114.754 Million cell updates/sec

Title: US-08-852-495C-1\_COPY\_1\_29000  
Perfect score: 29000  
Sequence: 1 CACACACACACACACACACA.....CCAGGCTAGAGTCAGTGGC 29000

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 374504

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/lna/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/lna/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/lna/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/lna/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/lna/6\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/lna/PCTUS\_COMB.seq:\*  
7: /cgn2\_6/ptodata/1/lna/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	71.4	0.2	105	4	US-08-481-658B-65
2	71.4	0.2	105	4	US-08-477-504A-65
3	71.4	0.2	105	4	US-08-486-756A-65
4	71.4	0.2	105	4	US-08-485-862B-65
5	71.4	0.2	105	5	US-08-787-739-65
6	69.6	0.2	105	4	US-08-481-658B-65
7	69.6	0.2	105	4	US-08-477-504A-65
8	69.6	0.2	105	4	US-08-486-756A-65
9	69.6	0.2	105	4	US-08-485-862B-65
10	69.6	0.2	105	5	US-08-787-739-65
11	58.6	0.2	78	3	US-08-454-557C-70
12	58.6	0.2	78	4	US-08-340-426D-70
13	58.6	0.2	78	4	US-08-450-673C-70
14	58.6	0.2	78	6	PCT-US95-17111A-70
15	57.2	0.2	78	3	US-08-454-557C-91
16	57.2	0.2	78	3	US-08-454-557C-70
17	57.2	0.2	78	4	US-08-340-426D-70
18	57.2	0.2	78	4	US-08-450-673C-70
19	55.8	0.2	84	3	US-08-454-557C-91
20	55.8	0.2	84	4	US-08-340-426D-91
21	55.8	0.2	84	4	US-08-450-673C-91
22	55.8	0.2	84	6	PCT-US95-17111A-91
23	54.8	0.2	76	3	US-08-454-557C-69
24	54.8	0.2	76	4	US-08-340-426D-69
25	54.8	0.2	76	4	US-08-450-673C-69
26	54.8	0.2	76	6	PCT-US95-17111A-69
27	55	0.2	84	3	US-08-454-557C-91

Sequence 91, Appl  
Sequence 91, Appl  
Sequence 91, Appl  
Sequence 60, Appl  
Sequence 60, Appl  
Sequence 60, Appl  
Sequence 60, Appl  
Sequence 60, Appl  
Sequence 92, Appl  
Sequence 92, Appl  
Sequence 92, Appl  
Sequence 36, Appl  
Sequence 57, Appl  
Sequence 57, Appl  
Sequence 57, Appl  
Sequence 60, Appl  
Sequence 60, Appl

ALIGNMENTS

RESULT 1  
US-08-481-658B-65  
; Sequence 65, Application US/08481658B  
; Patent No 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,658B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.3E  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-481-658B-65

Query Match 0.2%; Score 71.4; DB 4; Length 105;  
Best Local Similarity 80.0%; Pred. No. 5.2e-07;

**Matches** 84; **Conservative** 0; **Mismatches** 21; **Indels** 0; **Gaps** 0;

Qy	3720	TTTTTTTTTTATTTAGTAGAGATGGGGTTTACCCGTGTAGCCAGAATGGTCTCGATCTC	3779
Db	1	TTTTTTACATCTTTAGTAGAGACAGGGTTTCACCATATTGCCCGGCTGCTCTCAAACTC	60

**Qy**    3780 TTGACCTTCTGATCCGCCTTGCCCTTGGCTTCCCAGGTGCTGGGAT 3824  
         ||||| ||||| || |||| | ||||| ||||| |||||  
**Dβ**    61 CTGACCTTTGTGATCCACGAGCCTCGGCCCTCCCAGGTGCTGGGAT 105

## RESULT 2

US-08-477-504A-65  
; Sequence 65, Application US/08477504A

: FACILE NO. 3572333  
 : GENERAL INFORMATION:

```

1  APPLICANT: Zavada, Jan
2  APPLICANT: Pastorekova, Silvia
3  APPLICANT: Pastorek, Jaromir
4  TITLE OF INVENTION: MN Gene and Protein
5  NUMBER OF SEQUENCES: 86
6  CORRESPONDENCE ADDRESS:
7

```

ADDRESS: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California

```

; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

```

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,504A
; FILING DATE: 07-JUN-1995
;

```

```

: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/260,190
: FILING DATE: 15-JUN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Lauder, Leona L.
:

```

```

; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.3D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; INFORMATION FOR SEQ ID NO: 65:

```

Query Match 0.28; Score 71.4; DB 4; Length 105;  
Best Local Similarity 80.0%; Pred. No. 5.2e-07;  
Matches 84: Conservative 0; Mismatches 21: Indels

QY 3720 TTTT TTTT TTTT ATTT TAGT AGAG ATGGG GTTTCACCG TTAGCCAG AATGGTCTCGATCTC 3779

Db 1 TTTT TTTT ACATCTTT TAGT AGAGACAGGG TTTTCA CCAATATTTGGCCAGGCTGCTCAACTC 60

Qy	3780	TTGACCTTCTGATCCGCCCTGCCTTGGCTTCCCAAAGTGCCTGGGAT	3824
Dh	61	CTGACCTTGTGATCCACCAAGCTTGGCTTCCCAAAGTGCCTGGGAT	105

### RESULT 3

US-08-486-756A-65  
; Sequence 65, Application US/08486756A

; Patent No. 5981711  
; GENERAL INFORMATION:

```

;
; APPLICANT: Zavada, Jan
;
; APPLICANT: Pastorekova, Silvia
;
; APPLICANT: Pastorek, Jaromir
;
; TITLE OF INVENTION: MN Gene and Protein
;
; NUMBER OF SEQUENCES: 86
;
; CORRESPONDENCE ADDRESS:
;

```

ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920

```

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90
91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150
151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210
211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270
271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330
331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390
391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420
421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450
451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510
511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570
571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600
601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630
631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660
661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690
691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720
721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750
751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780
781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810
811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840
841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870
871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900
901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930
931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960
961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990
991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020
1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 
```

; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/486,756A  
 ; FILING DATE: 07-JUN-1995  
 ; CLASSIFICATION: 424

1 CONTRIBUTION: 124  
 2 PRIOR APPLICATION DATA:  
 3 APPLICATION NUMBER: US 08/260,190  
 4 FILING DATE: 15-JUN-1994  
 5 ATTORNEY/AGENT INFORMATION:  
 6 NAME: Lauder, Leona L.  
 7 REGISTRATION NUMBER: 20 963

```

:
:
: REGISTRATION NUMBER: 50,86
: REFERENCE/DOCKET NUMBER: D
: TELECOMMUNICATION INFORMATION
: TELEPHONE: 415-435-2034
: TELEFAX: 415-435-0727
: INFORMATION FOR SEQ ID NO: 65:
:

```

Query Match	0.2%	Score	71.4;	DB	4;	Length	105;
Best Local Similarity	80.0%;	Pred. NO.	5.2e-07;				
Matches	84;	Conservative	0;	Mismatches	21;	Indels	0;
						Gaps	0;

QY 3720 TTTT TTTT TTTT TTTT TAGTAGAGATGGGGTTTTCACCGTGTACCCAGAAATGGTCTCGATCTC 3779

db 1 TTTT TTTT TTTT TTTT TAGTAGAGAGACAGGAGTTCACCATATTTGCCAGGCTGCTCAAACTC 50

QY 3780 TTGACCTTCTGATCCGCCCTGCCTTGCCCTTGGCTTCCCCAAAGTGCTGGGAT 3824  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 61 CTGACCTTGTGATCCACACAGCCTCGGCCCTCCCCAAGTGCTGGGAT 105

## RESULT

US-08-485-862B-65  
; Sequence 65, Application US/08485862B

Patent No. 5989838

```

; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
;

```

;  
; ADDRESS: Leona L. Lauder  
; STREET: 6 Mariposa Court

CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,862B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-485-862B-65

Query Match 0.2%; Score 71.4; DB 4; Length 105;  
Best Local Similarity 80.0%; Pred. No. 5.2e-07;  
Matches 84; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3720 TTTTATTTTATTTAGTAGAGATGGGTTTACCGTGTAGCCAGAAATGGTCTCGATCTC 3779  
Db 1 TTTTATACATCTTTAGTAGACAGAGGTTTACCATATTTGGCAGGCTGCTCTCAAATC 60

QY 3780 TTGACCTTCTGATCCCGCTTGGCTTCCCAAAGTCTGGGAT 3824  
Db 61 CTGACCTTGTGTATCCACAGCCTCGCCCTCCCAAAGTCTGGGAT 105

RESULT 5  
US-08-787-739-65  
; Sequence 65, Application US/08/787739  
; Patent No. 6027887  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 96  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 369 Pine Street, Suite 610  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/787,739  
FILING DATE: 24-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,049  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/486,756  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/481,658  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,862  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,863  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/487,077  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-981-0332  
TELEFAX: 415-981-0332  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-787-739-65

Query Match 0.2%; Score 71.4; DB 5; Length 105;  
Best Local Similarity 80.0%; Pred. No. 5.2e-07;  
Matches 84; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3720 TTTTATTTTATTTAGTAGAGATGGGTTTACCGTGTAGCCAGAAATGGTCTCGATCTC 3779  
Db 1 TTTTATACATCTTTAGTAGACAGAGGTTTACCATATTTGGCAGGCTGCTCTCAAATC 60

QY 3780 TTGACCTTCTGATCCCGCTTGGCTTCCCAAAGTCTGGGAT 3824  
Db 61 CTGACCTTGTGTATCCACAGCCTCGCCCTCCCAAAGTCTGGGAT 105

RESULT 6  
US-08-481-658B-65/c  
; Sequence 65, Application US/08/481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/481.658B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-481-658B-65

Query Match 0.2%; Score 69.6; DB 4; Length 105;  
Best Local Similarity 85.6%; Pred. No. 1.3e-06;  
Matches 89; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

Qy 1395 ATCTCAGACACTTTGGGAGGCTGAGG-GCACAGATCAGAGGTGGGAGTTTGAGACCAGC 1453  
||| ||||||||||||||| ||||| ||||| ||||||||||| |||||  
Db 105 ATCCAGCAGCACTTTGGGAGCGCGAGCTGGTGGATCACAAGGTCAGGAGTTTGAGAGCAGC 46

Qy 1454 CTGGCCAATATGGCAACCCCTGCTCTACTAAATAACAAAA 1497  
||||| ||||||| ||||||| ||||||| ||||||| || |||||  
Db 45 CTGGCCAATATGGTGAACCCCTGCTCTACTAAAGATGTAANA 2

RESULT 7  
US-08-477-504A-65/c  
Sequence 65, Application US/08477504A  
Patent No. 5972353  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477.504A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:

NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-477-504A-65

Query Match 0.2%; Score 69.6; DB 4; Length 105;  
Best Local Similarity 85.6%; Pred. No. 1.3e-06;  
Matches 89; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

Qy 1395 ATCTCAGACACTTTGGGAGGCTGAGG-GCACAGATCAGAGGTGGGAGTTTGAGACCAGC 1453  
||| ||||||||||||||| ||||| ||||| ||||||||||| |||||  
Db 105 ATCCAGCAGCACTTTGGGAGCGCGAGCTGGTGGATCACAAGGTCAGGAGTTTGAGAGCAGC 46

Qy 1454 CTGGCCAATATGGCAACCCCTGCTCTACTAAATAACAAAA 1497  
||||| ||||||| ||||||| ||||||| ||||||| || |||||  
Db 45 CTGGCCAATATGGTGAACCCCTGCTCTACTAAAGATGTAANA 2

RESULT 8  
US-08-486-756A-65/c  
Sequence 65, Application US/08486756A  
Patent No. 5981711  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486.756A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-486-756A-65

Query Match 0.2%; Score 69.6; DB 4; Length 105;  
Best Local Similarity 85.6%; Pred. No. 1.3e-06;  
Matches 89; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 1395 ATCTCAGCACTTTGGAGGCTGAGG-GCACAGATCACAGGTCGGGAGTTTGAGACCAGC 1453  
DB 105 ATCCAGCACTTTGGAGGCGCCGAGGCTGGTGGATCACAGGTCAGGAGTTTGAGACCAGC 46

QY 1454 CTGGCCAATATGCGGAACCCCTCTCTCTACTAAATATACAAAA 1497  
DB 45 CTGGCCAATATGCGGAACCCCTCTCTCTACTAAAGATGTAAAAA 2

## RESULT 9

US-08-485-862B-65/C  
Sequence 65, Application US/08485862B  
Patent No. 5989838

GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA

ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION NUMBER: US/08/485,862B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727

INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-485-862B-65

Query Match 0.2%; Score 69.6; DB 4; Length 105;  
Best Local Similarity 85.6%; Pred. No. 1.3e-06;  
Matches 89; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 1395 ATCTCAGCACTTTGGAGGCTGAGG-GCACAGATCACAGGTCGGGAGTTTGAGACCAGC 1453  
DB 105 ATCCAGCACTTTGGAGGCGCCGAGGCTGGTGGATCACAGGTCAGGAGTTTGAGACCAGC 46  
QY 1454 CTGGCCAATATGCGGAACCCCTCTCTCTACTAAATATACAAAA 1497  
DB 45 CTGGCCAATATGCGGAACCCCTCTCTCTACTAAAGATGTAAAAA 2

## RESULT 10

US-08-787-739-65/C  
Sequence 65, Application US/08787739  
Patent No. 6027387

GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 96  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 369 Pine Street, Suite 610  
CITY: San Francisco  
STATE: California  
COUNTRY: USA

ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION NUMBER: US/08/787,739  
FILING DATE: 24-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,049  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/486,756  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/481,658  
FILING DATE: 07-JUN-1995

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,862  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,863  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/487,077  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-981-2034  
TELEFAX: 415-981-0332

INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-787-739-65

```
Query Match          0.2%; Score 69.6; DB 5; Length 105;
Best Local Similarity 85.6%; Pred. No. 1.3e-06;
Matches 89; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

Qy 1395 ATCTCAGCATTGGGAGGTGAGG-GCACAGATCACAGGTGGGAGTTTGACAGCAGC 1453
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 105 ATCCAGCAGCATTGGGAGCGGAGGCTGGTGGATCACAGGTTCAGAGTTTGAGAGCAGC 46

Qy 1454 CTGGCCAATATGCGAAACCCCTGCTCTACTAAATAACAAAA 1497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 45 CTGGCCAATATGCTGAACCCCTGCTCTACTAAAGATGTAATAA 2

RESULT 11
US-08-454-557C-70
; Sequence 70, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 30-MAY-1995
; APPLICATION NUMBER: US/08/454,557C
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
US-08-454-557C-70

Query Match          0.2%; Score 58.6; DB 3; Length 78;
Best Local Similarity 87.7%; Pred. No. 0.00029;
Matches 64; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 2840 GCCCAGCTAATTTTGTATTTTAGTAGAGATGGGGTTTCTACTATGTGTGGCCAGGCTAGT 2899
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6 GCCCAGCTAATTTTGTATTTTAGTAGAGATGGGGTTTCTCTCCATGTTTCATCAGGCTGTT 65

Qy 2900 TTGGAACCTCTCA 2912
    | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66 GTCGAACCTCTCA 78

RESULT 12
US-08-340-426D-70
; Sequence 70, Application US/08340426D
; Patent No. 5948634
```

```
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 14-NOV-1994
; APPLICATION NUMBER: US/08/340,426D
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
US-08-340-426D-70

Query Match          0.2%; Score 58.6; DB 4; Length 78;
Best Local Similarity 87.7%; Pred. No. 0.00029;
Matches 64; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 2840 GCCCAGCTAATTTTGTATTTTAGTAGAGATGGGGTTTCTACTATGTGTGGCCAGGCTAGT 2899
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6 GCCCAGCTAATTTTGTATTTTAGTAGAGATGGGGTTTCTCCATGTTTCATCAGGCTGTT 65

Qy 2900 TTGGAACCTCTCA 2912
    | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66 GTCGAACCTCTCA 78

RESULT 13
US-08-450-673C-70
; Sequence 70, Application US/08450673C
; Patent No. 5948888
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```



```
; TOPOLOGY: both
PCT-US95-17IIIA-70

Query Match          0.2%   Score 58.6; DB 6; Length 78;
Best Local Similarity 87.7%; Pred No. 0.00029;
Matches 64; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2840 GCCCAGCTAAATTTTGTGTAATTTTTAGTAGAGATGGGGTTTCACATATGTTGGCCAGGCTAGT 2899
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    6 GCCCAGCTAAATTTGTGTAATTTTTAGTAGAGATGGGGTTTCACATATGTTGCATCAGGCTGCT 65

QY 2900 TTGGAACCTCCTGA 2912
      | |||||||||
Db    66 GTCGAACCTCCTGA 78

RESULT 15
US-08-454-557C-70/C
; Sequence 70, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: Of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
```

```

? COUNTRY: U.S.A.
? ZIP: 20005-3934
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/454.557C
? FILING DATE: 30-MAY-1995
? CLASSIFICATION: 51A
? ATTORNEY/AGENT INFORMATION:
? NAME: Ludwig, Steven R.
? REGISTRATION NUMBER: 36,203
? REFERENCE/DOCKET NUMBER: 0609.3840003
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 371-2600
? TELEFAX: (202) 371-2540
? INFORMATION FOR SEQ ID NO: 70:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 78 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: both
? TOPOLOGY: both
?
? US-08-454-557C-70

```

```

Query Match          0.2%;   Score 57.2;   DB 3;   Length 78;
Best Local Similarity 83.3%;   Pred. No. 0.00059;
Matches    65;   Conservative    0;   Mismatches    13;   Indels    0;   Gaps    0

```

```

QY  1435 TCGGGAGTTTCAGACCGCTGGCCAATATGCGGAACCCCTGTCTCTACTATAAAATACAA 1494
      |||||
Db   78 TCAGATTTCAGCACACCAGCGTATGAACATGGAGAACCCCATCTCTACTATAAAATACAA 19
      |||||
QY  1495 AAATTAGCTGGGCATGGT 1512
      |||||
Db   18 ATATTAGCTGGCGTGTT 1

```

Search completed: June 14, 2000, 20:28:41  
Job time: 27251 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2000, 20:22:29 ; Search time 17971.3 Seconds  
(without alignments)  
-1569.830 Million cell updates/sec

Title: US-08-852-495C-1\_COPY\_28000\_57000  
Perfect score: 29001  
Sequence: 1 ATATCAACAAAAACACACAT.....TTAGCAGCACACAGGTAGGTT 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 370290

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl.\*

1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_on.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pl1.\*  
8: gb\_pl2.\*  
9: gb\_pr1.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: gb\_ro.\*  
13: gb\_sts.\*  
14: gb\_sy.\*  
15: gb\_un.\*  
16: gb\_vi.\*  
17: em\_fun.\*  
18: em\_hum1.\*  
19: em\_hum2.\*  
20: em\_in.\*  
21: em\_on.\*  
22: em\_or.\*  
23: em\_ov.\*  
24: em\_pat.\*  
25: em\_ph.\*  
26: em\_pl.\*  
27: em\_ro.\*  
28: em\_sts.\*  
29: em\_sy.\*  
30: em\_un.\*  
31: em\_vi.\*  
32: gb\_htg1.\*  
33: gb\_htg2.\*  
34: gb\_in1.\*  
35: gb\_in2.\*  
36: em\_ba1.\*  
37: em\_ba2.\*  
38: em\_hum3.\*  
39: em\_hum4.\*  
40: gb\_pr4.\*  
41: gb\_htg3.\*  
42: gb\_htg4.\*  
43: gb\_htg5.\*  
44: gb\_htg6.\*

45: gb\_htg7.\*  
46: em\_htg1.\*  
47: em\_htg2.\*  
48: em\_htg3.\*  
49: em\_hum5.\*  
50: gb\_pl3.\*  
51: gb\_pr5.\*  
52: gb\_htg8.\*  
53: gb\_htg9.\*  
54: gb\_htg10.\*  
55: gb\_htg11.\*  
56: gb\_htg12.\*  
57: gb\_htg13.\*  
58: gb\_htg14.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	94.4	0.3	96	13	G31304	G31304 sy899g1-19
2	91.6	0.3	108	10	HSLDLRN2	X05250 Human LDL-r
3	90.4	0.3	108	10	HSLDLRN2	X05250 Human LDL-r
4	88	0.3	108	11	HSU67803	U67803 Human small
5	87	0.3	107	9	HUMALCE162	M87924 Human carc
6	87	0.3	108	11	HSU67804	U67804 Human small
7	85.4	0.3	103	9	HUMALCE221	M87896 Human carc
8	84	0.3	103	9	HUMALCE221	M87896 Human carc
9	84.2	0.3	108	10	HSLDLRD1	X05249 Human LDL-r
10	84.2	0.3	108	10	HSLDLRD2	X05251 Human LDL-r
11	83	0.3	108	10	HSLDLRD1	X05249 Human LDL-r
12	83	0.3	108	10	HSLDLRD2	X05251 Human LDL-r
13	81.4	0.3	108	10	HSLDLI12	X05248 Human LDL-r
14	79	0.3	108	11	HSU67808	U67808 Human small
15	78.8	0.3	110	9	HUMALCE43	M87900 Human carc
16	78.2	0.3	110	11	HSU67807	U67807 Human small
17	77.4	0.3	90	9	HUMLDLRFL	X03555 Human low d
18	77.6	0.3	103	13	HS8IC8R	X57789 Human sequ
19	77.6	0.3	104	9	HUMALCE272	M87899 Human carc
20	77.4	0.3	107	9	HUMALCE162	M87924 Human carc
21	77.2	0.3	108	10	HSLDLI12	X05248 Human LDL-r
22	76.2	0.3	108	11	HSU67803	U67803 Human small
23	75.4	0.3	97	9	HUMLDLRDJ	M14179 Human faml
24	75.6	0.3	103	13	HS8IC8R	X57789 Human sequ
25	75	0.3	110	11	HSU67807	U67807 Human small
26	73.8	0.3	97	9	HUMLDLRA1	M14178 Human low d
27	73.6	0.3	97	9	HUMLDLRA2	M14180 Human low d
28	73	0.3	107	11	HSU67806	U67806 Human small
29	73.2	0.3	110	9	HUMALCE43	M87900 Human carc
30	72.6	0.3	100	9	HUMALNSA	D45223 Human GALNS
31	72.6	0.3	108	9	HUMDI003M5	D16965 Human HepG2
32	72.2	0.2	108	9	HUMDI003M5	D16965 Human HepG2
33	71	0.2	104	9	HUMALCE272	M87899 Human carc
34	71	0.2	107	11	HSU67806	U67806 Human small
35	70.6	0.2	101	10	S79560	S79560 HRX (intron
36	69.8	0.2	108	11	HSU67808	U67808 Human small
37	69.2	0.2	91	13	HUMUT8164A	L30244 Human STS U
38	69.2	0.2	99	13	HUMUT7692A	L30306 Human STS U
39	69.2	0.2	100	13	HUMUT931A	L31299 Human STS U
40	69.4	0.2	108	13	G43535	G43535 Human STS U
41	68.8	0.2	80	9	HUMBRKFAE	M36135 Human alpha
42	68.8	0.2	108	13	G32614	G32614 A009K21 Hum
43	67.6	0.2	91	13	HUMUT8164A	L30244 Human STS U
44	67.8	0.2	100	9	HUMALNSA	D45223 Human GALNS
45	67.6	0.2	107	13	G32919	G32919 A009W27 Hum

ALIGNMENTS



## ORIGIN

Query Match 0.3%; Score 90.4; DB 10; Length 108;  
Best Local Similarity 89.8%; Pred. No. 2.5e-07;  
Matches 97; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 7174 AAAAAATTAGCGGTGATGGCCGTCGCTGTAGTCTCAGCCACTTGGGAGGCTGAG 7233

Db 1 AAAAAATTAGCGAGCGGTGGTGGCAGGTGCTGTATCCAGCTACTCGGAGGCTGAG 60

QY 7234 GCAGGAAAAATGCTTGAACCCAGGAGGAGGTTGCTCAGTGAGCCGAG 7281

Db 61 GCAGGAGAAATGCTTGAACCCAGGAGGAGGTTGCTCAGTGAGCCGAG 108

RESULT 4  
HSU67803/c  
LOCUS HSU67803 108 bp RNA PRI 01-AUG-1997  
DEFINITION Human small cytoplasmic Alu transcript.  
ACCESSION U67803  
VERSION U67803.1 GI:2289917  
KEYWORDS Alu.

SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 108)  
AUTHORS Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.  
TITLE cDNAs derived from primary and small cytoplasmic Alu (sAlu)  
transcripts

JOURNAL J. Mol. Biol. 271 (2), 222-234 (1997)  
MEDLINE 97415756  
REFERENCE 2 (bases 1 to 108)  
AUTHORS Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.  
TITLE Direct Submission  
JOURNAL Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The Children's Hospital of Philadelphia, 1004F Abramson Research Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA

## FEATURES

source  
1. .108  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="TscAlu2"  
repeat\_region 1. .108  
/note="sAlu"  
/rpt\_family="Alu"  
/rpt\_type="dispersed"

BASE COUNT 23 a 39 c 30 g 16 t

Query Match 0.3%; Score 88; DB 11; Length 108;  
Best Local Similarity 94.8%; Pred. No. 7e-07;  
Matches 91; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 10907 TAGAGATGGGTTTACCATTAGCCAGGATGCTCGATCTCCACCTCGTGATCCA 10966

Db 96 TAGAGAGGGGTTTACCATTAGCCAGGATGCTCGATCTCCACCTCGTGATCCG 37

QY 10967 CCGCTTTGGCCCTCCCAAGTCTGGGATTACAGGC 11002

Db 36 CCGCCCTCGCCCTCCCAAGTCTGGGATTACAGGC 1

RESULT 5  
HUMALCE162/c  
LOCUS HUMALCE162 107 bp ss-RNA PRI 15-APR-1994  
DEFINITION Human carcinoma cell-derived Alu RNA transcript, clone CE162.  
ACCESSION M87924  
VERSION M87924.1 GI:174871  
KEYWORDS Alu repeat.  
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 107)  
AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.  
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of post-transcriptional selection of master sequences  
J. Mol. Biol. (1992) In press

## JOURNAL

Location/Qualifiers

## FEATURES

source  
1. .107  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Ntera2D1"  
/dev\_stage="embryo"  
/sex="male"  
/tissue\_type="carcinoma"

BASE COUNT 28 a 30 c 35 g 14 t

Query Match 0.3%; Score 87; DB 9; Length 107;  
Best Local Similarity 90.3%; Pred. No. 1.1e-06;  
Matches 93; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 10736 TTTTTCAGATGAGTCTCAGTCTGTACCTAGGCTGAGTGGCGCAACTCGC 10795

Db 107 TTTTTCAGATGAGTCTCAGTCTGTACCTAGGCTGAGTGGCGCAACTCGC 48

QY 10796 TCACTGAAGCTTGCCTCTGGTTCATGCCATTCCTCTGCC 10838

Db 47 TCACTGAAGCTTGCCTCTGGTTCATGCCATTCCTCTGCC 5

## RESULT 6

HSU67804/c

LOCUS HSU67804 108 bp RNA PRI 01-AUG-1997  
DEFINITION Human small cytoplasmic Alu transcript.  
ACCESSION U67804  
VERSION U67804.1 GI:2289918  
KEYWORDS Alu.

## SOURCE

human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 108)  
AUTHORS Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.  
TITLE cDNAs derived from primary and small cytoplasmic Alu (sAlu)  
transcripts

JOURNAL J. Mol. Biol. 271 (2), 222-234 (1997)  
MEDLINE 97415756

## REFERENCE

2 (bases 1 to 108)

AUTHORS Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.

TITLE Direct Submission

JOURNAL Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The Children's Hospital of Philadelphia, 1004F Abramson Research Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA

## FEATURES

source  
1. .108  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="TscAlu3"  
repeat\_region 1. .108  
/note="sAlu"  
/rpt\_family="Alu"  
/rpt\_type="dispersed"

BASE COUNT 26 a 38 c 26 g 18 t

Query Match 0.3%; Score 87; DB 11; Length 108;  
Best Local Similarity 94.7%; Pred. No. 1.1e-06;  
Matches 90; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 10908 AGAGATGGGGTTTACCATGTTAGCCAGATGGTCTCGATCTCTCGACTCGTGATCCAC 10967  
 Db 95 AAAGACGGGGTTTACCATGTTAGCCAGATGGTCTCGATCTCTCGACTCGTGATCCAC 36  
 Qy 10968 CGGCTTTGGCCTCCCAAAAGTCTGGGATTACAGGC 11002  
 Db 35 CGGCTTTGGCCTCCCAAAAGTCTGGGATTACAGGC 1

RESULT 7  
 HUMALCE221 103 bp ss-RNA PRI 15-APR-1994  
 LOCUS Human carcinoma cell-derived Alu RNA transcript, clone CE221.  
 DEFINITION M87896.  
 ACCESSION M87896.  
 VERSION M87896.1 GI:174874  
 KEYWORDS Alu repeat.  
 SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 103)  
 AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.  
 TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of  
 post-transcriptional selection of master sequences  
 JOURNAL J. Mol. Biol. (1992) In press  
 FEATURES Location/Qualifiers  
 source 1..103  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_line="Ntera2D1"  
 /dev\_stage="embryo"  
 /sex="male"  
 /tissue\_type="carcinoma"

BASE COUNT 25 a 27 c 33 g 18 t  
 ORIGIN

Query Match 0.3%; Score 85.4; DB 9; Length 103;  
 Best Local Similarity 89.3%; Pred. No. 2.2e-06;  
 Matches 92; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
 Qy 20749 ATGCACCTGTATTCAGCTACTCGGGAGCTGACAGAGAGATCGCTTGAACCTGGGA 20808  
 Db 1 ATCTCGCTGTATTCAGCTACTCGGGAGCTGACAGAGAGATCGCTTGAACCTGGGA 60  
 Qy 20809 GCGGAGGTGGTGGAGCGAGATCGCATTCCTCCAG 20851  
 Db 61 GCGGAGGTGGTGGAGCGAGATCGCATTCCTCCAG 103

RESULT 8  
 HUMALCE221/c 103 bp ss-RNA PRI 15-APR-1994  
 LOCUS Human carcinoma cell-derived Alu RNA transcript, clone CE221.  
 DEFINITION M87896  
 ACCESSION M87896  
 VERSION M87896.1 GI:174874  
 KEYWORDS Alu repeat.  
 SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 103)  
 AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.  
 TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of  
 post-transcriptional selection of master sequences  
 JOURNAL J. Mol. Biol. (1992) In press  
 FEATURES Location/Qualifiers  
 source 1..103  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_line="Ntera2D1"  
 /dev\_stage="embryo"  
 /sex="male"

BASE COUNT 25 a 27 c 33 g 18 t  
 ORIGIN

Query Match 0.3%; Score 84; DB 9; Length 103;  
 Best Local Similarity 90.0%; Pred. No. 4e-06;  
 Matches 90; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 5440 CTGGAGTCAATGGCGGATCTCGCTCACCACACCTCTACTCCAGGTTCAAGCAAT 5499  
 Db 103 CTGGAGTCAATGGCAGCATCTCGGCTCACTGCACCTCCCGGTTCAAGCGAT 44  
 Qy 5500 TCTCTGCTGCTCGCTCCGAGTAGCTGGGATTACAGGCA 5539  
 Db 43 TCTCTGCTGCTGCTCCGCTAGTCCGCTGCTAGTGGGATTACAGGCA 4

RESULT 9

HSLDLRD1

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

\*source: hypercholesterol aemia

See X05248 for corresponding normal gene sequence

In the defective LDL-receptor gene the deletion occurred between two

alu-repetitive sequences, that are in the same direction, the

deletion eliminates exons 13 and 14 and changes the reading frame

of the resulting spliced mRNA.

Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES

source 1..108

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/cell\_type="blood leukocytes from a patient with familial"

misc\_feature 1..108

/note="deletion junction region intron 12/ intron 15"

BASE COUNT 20 a 40 c 20 g 28 t

ORIGIN

Query Match 0.3%; Score 84.2; DB 10; Length 108;

Best Local Similarity 87.6%; Pred. No. 3.6e-06;

Matches 92; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 10791 TCGGCTACTCGAAGCTCTGCTCTGGGTTTCATGCCATTCCTCGCTCAGCCTCCGA 10850

Db 2 TCGGCTACCACACCTCTGCTCTGGGTTTCATGCCATTCCTCGCTCAGCCTCCGA 61

Qy 10851 GTAGCTGGGACTACAGGCGTCTGCCACCACGCCAGCTAAATTTT 10895

Db 62 GTAGCTGGGATTACAGGCGTCTGCCACCACGCCAGCTAAATTTT 106

RESULT 10

HSLDLRD2/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

\*source: hypercholesterol aemia

See X05248 for corresponding normal gene sequence

In the defective LDL-receptor gene the deletion occurred between two

alu-repetitive sequences, that are in the same direction, the

deletion eliminates exons 13 and 14 and changes the reading frame

of the resulting spliced mRNA.

Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES

source 1..108

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/cell\_type="blood leukocytes from a patient with familial"

misc\_feature 1..108

/note="deletion junction region intron 12/ intron 15"

BASE COUNT 20 a 40 c 20 g 28 t

ORIGIN

Query Match 0.3%; Score 84.2; DB 10; Length 108;

Best Local Similarity 87.6%; Pred. No. 3.6e-06;

Matches 92; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 10791 TCGGCTACTCGAAGCTCTGCTCTGGGTTTCATGCCATTCCTCGCTCAGCCTCCGA 10850

Db 2 TCGGCTACCACACCTCTGCTCTGGGTTTCATGCCATTCCTCGCTCAGCCTCCGA 61

Qy 10851 GTAGCTGGGACTACAGGCGTCTGCCACCACGCCAGCTAAATTTT 10895

Db 62 GTAGCTGGGATTACAGGCGTCTGCCACCACGCCAGCTAAATTTT 106

RESULT 10

HSLDLRD2/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

\*source: hypercholesterol aemia

See X05248 for corresponding normal gene sequence

In the defective LDL-receptor gene the deletion occurred between two

alu-repetitive sequences, that are in the same direction, the

deletion eliminates exons 13 and 14 and changes the reading frame

of the resulting spliced mRNA.

Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES

source 1..108

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/cell\_type="blood leukocytes from a patient with familial"

misc\_feature 1..108

/note="deletion junction region intron 12/ intron 15"

BASE COUNT 20 a 40 c 20 g 28 t

ORIGIN

Query Match 0.3%; Score 84.2; DB 10; Length 108;

Best Local Similarity 87.6%; Pred. No. 3.6e-06;

Matches 92; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 10791 TCGGCTACTCGAAGCTCTGCTCTGGGTTTCATGCCATTCCTCGCTCAGCCTCCGA 10850

Db 2 TCGGCTACCACACCTCTGCTCTGGGTTTCATGCCATTCCTCGCTCAGCCTCCGA 61

Qy 10851 GTAGCTGGGACTACAGGCGTCTGCCACCACGCCAGCTAAATTTT 10895

Db 62 GTAGCTGGGATTACAGGCGTCTGCCACCACGCCAGCTAAATTTT 106

RESULT 10

HSLDLRD2/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

\*source: hypercholesterol aemia

See X05248 for corresponding normal gene sequence

In the defective LDL-receptor gene the deletion occurred between two

alu-repetitive sequences, that are in the same direction, the

deletion eliminates exons 13 and 14 and changes the reading frame

of the resulting spliced mRNA.

Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES

source 1..108

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/cell\_type="blood leukocytes from a patient with familial"

misc\_feature 1..108

/note="deletion junction region intron 12/ intron 15"

BASE COUNT 20 a 40 c 20 g 28 t

ORIGIN

Query Match 0.3%; Score 84.2; DB 10; Length 108;

Best Local Similarity 87.6%; Pred. No. 3.6e-06;

Matches 92; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 10791 TCGGCTACTCGAAGCTCTGCTCTGGGTTTCATGCCATTCCTCGCTCAGCCTCCGA 10850

Db 2 TCGGCTACCACACCTCTGCTCTGGGTTTCATGCCATTCCTCGCTCAGCCTCCGA 61

Qy 10851 GTAGCTGGGACTACAGGCGTCTGCCACCACGCCAGCTAAATTTT 10895

Db 62 GTAGCTGGGATTACAGGCGTCTGCCACCACGCCAGCTAAATTTT 106

RESULT 10

HSLDLRD2/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

\*source: hypercholesterol aemia

See X05248 for corresponding normal gene sequence

In the defective LDL-receptor gene the deletion occurred between two

alu-repetitive sequences, that are in the same direction, the

deletion eliminates exons 13 and 14 and changes the reading frame

of the resulting spliced mRNA.

Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES

source 1..108

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/cell\_type="blood leukocytes from a patient with familial"

misc\_feature 1..108

/note="deletion junction region intron 12/ intron 15"

BASE COUNT 20 a 40 c 20 g 28 t

ORIGIN

Query Match 0.3%; Score 84.2; DB 10; Length 108;

Best Local Similarity 87.6%; Pred. No. 3.6e-06;

Matches 92; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 10791 TCGGCTACTCGAAGCTCTGCTCTGGGTTTCATGCCATTCCTCGCTCAGCCTCCGA 10850

Db 2 TCGGCTACCACACCTCTGCTCTGGGTTTCATGCCATTCCTCGCTCAGCCTCCGA 61

Qy 10851 GTAGCTGGGACTACAGGCGTCTGCCACCACGCCAGCTAAATTTT 10895

Db 62 GTAGCTGGGATTACAGGCGTCTGCCACCACGCCAGCTAAATTTT 106

RESULT 10

HSLDLRD2/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

\*source: hypercholesterol aemia

See X05248 for corresponding normal gene sequence

In the defective LDL-receptor gene the deletion occurred between two

alu-repetitive sequences, that are in the same direction, the

deletion eliminates exons 13 and 14 and changes the reading frame

of the resulting spliced mRNA.

Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES

source 1..108

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/cell\_type="blood leukocytes from a patient with familial"

misc\_feature 1..108







---



Result No.	Score	Query		DB	ID	Description
		Match	Length			
C 1	70	0.2	108	1	X12095	Human biallelic po
C 2	67.2	0.2	100	1	T24892	Human gene signatu
C 3	66	0.2	103	1	T20927	Human gene signatu
C 4	65.2	0.2	108	1	X12095	Human biallelic po
C 5	64	0.2	100	1	X12086	Human biallelic po
C 6	63.6	0.2	100	1	X12087	Human biallelic po
C 7	63.6	0.2	100	1	X12085	Human biallelic po
C 8	62.4	0.2	100	1	T24892	Human gene signatu
C 9	62.6	0.2	103	1	T26213	Human gene signatu
C 10	62.2	0.2	108	1	T26828	Human gene signatu
C 11	60.2	0.2	103	1	T20927	Human gene signatu
C 12	59.4	0.2	91	1	T25854	Human gene signatu
C 13	59.4	0.2	108	1	T26828	Human gene signatu
C 14	58.6	0.2	108	1	T25009	Human gene signatu
C 15	57.8	0.2	108	1	T25009	Human gene signatu
C 16	57.2	0.2	91	1	T25854	Human gene signatu
C 17	55.6	0.2	103	1	T26213	Human gene signatu
C 18	55.2	0.2	87	1	T21566	Human gene signatu
C 19	55.4	0.2	110	1	T25260	Human gene signatu
C 20	55	0.2	93	1	T22572	Human gene signatu
C 21	54.2	0.2	95	1	T23131	Human gene signatu
C 22	54	0.2	99	1	T20931	Human gene signatu
C 23	53.4	0.2	87	1	T21566	Human gene signatu
C 24	53.2	0.2	93	1	T22572	Human gene signatu
C 25	53.4	0.2	97	1	T26728	Human gene signatu
C 26	53	0.2	109	1	T23895	Human gene signatu
C 27	52.4	0.2	70	1	N60231	Normal chromosome
C 28	52.6	0.2	93	1	T24259	Human gene signatu
C 29	52.2	0.2	100	1	X12085	Human biallelic po
C 30	52.4	0.2	110	1	T25260	Human gene signatu
C 31	52	0.2	69	1	T29016	Probe to internal
C 32	52	0.2	84	1	T25848	Human gene signatu
C 33	52	0.2	93	1	T25688	Human gene signatu
C 34	52	0.2	93	1	T25688	Human gene signatu

```
RESULT 2
T24892/c
ID T24892 standard; cDNA to mRNA; 100 BP.
AC T24892;
DE 05-NOV-1996 (first entry)
DE Human gene signature HUMGS06998.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 93-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1720; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 100 BP; 28 A; 22 C; 25 G; 22 T;

Query Match 0.2%; Score 67.2; DB 1; Length 100;
Best Local Similarity 78.8%; Pred. No. 0.041;
Matches 78; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 15777 TTTTCTTTTTCAGACAGAGTCTCACTATACACCCAGCTGGAGTGCAGTGGCAATC 15836
DB 100 TTGTGTTGTTTCAACAGAGTGTCACTCTGTCAACCCAGCGNGAGTGCAGNGTCAATC 41

QY 15837 TCAGCTCACTGCAACCTGCACCTCTCTGGGTTCAAGGGAT 15875
DB 40 TCAGCTNATTGCAAAATCTGCTCCAGCTTCAACGGAT 2

RESULT 3
T20927/c
ID T20927 standard; cDNA to mRNA; 103 BP.
AC T20927;
DE 24-JUL-1996 (first entry)
DE Human gene signature HUMGS02180.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 93-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
```

```
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 758-759; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 103 BP; 22 A; 27 C; 21 G; 31 T;

Query Match 0.2%; Score 66; DB 1; Length 103;
Best Local Similarity 78.0%; Pred. No. 0.062;
Matches 78; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 9894 CATGGTGAACCCCTGCTCTACTAAATAACAAAAATTAGCCGGCATGGTGCACGC 9953
DB 100 CATGGAGAATACTGTCCCTACTTNAATAATACAAATCAGTGGACATGGTGGCACAC 41

QY 9954 CTGTATATCCCACTACTTGGGAGCTGAGACAGGAGAAATC 9993
DB 40 CTGTAGCCACAGCTACTTGGGAGGTGGAAGTGGGAGGATC 1

RESULT 4
X12095
ID X12095 standard; DNA; 108 BP.
AC X12095;
DE 30-MAR-1999 (first entry)
DE Human biallelic polymorphic DNA fragment TIGR-A003M18a.
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medication;
KW treatment; marker; ss.
OS Homo sapiens.
PN W09820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; U20313.
PR 06-NOV-1996; US-030455.
PA (WHEED ) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
DR WPI; 98-286974/25.
PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
PS Claim 1; Page 219; 310pp; English.
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC X03121-X10268. The base occupying the polymorphic site is indicated by
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
```

CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases. 23 C; 28 G; 37 T;  
SQ Sequence 108 BP; 19 A; 23 C; 28 G; 37 T;

Query Match 0.2%; Score 65.2; DB 1; Length 108;  
Best Local Similarity 80.6%; Pred. No. 0.082;  
Matches 87; Conservative 1; Mismatches 19; Indels 1; Gaps 1;

QY 21852 TGTATTTTGTAGACAGCGGGTTTCCACATGTTGTCAGGCTGTCGGAACTCCTGAC 21911  
DB 1 TGTCTTTTGTAGAGATGAGGTTTTCCTRTGTGGCCAGGATGTCGGAACCTCTGAC 60

QY 21912 CTCAGGTGATCGCCACTCAGCTCCG-AAAAGTCGTGGGATTACAG 21958  
DB 61 TTCAGTGTGCTGCTGGCCCTCCAAAAGTCTGGGATTATAG 108

RESULT 5  
X12086/c  
ID X12086 standard; DNA; 100 BP.  
AC X12086;  
DE 30-MAR-1999 (first entry)  
KW Human biallelic polymorphic DNA fragment EST98276b.  
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;  
KW detection; phenotypic typing; characteristic; infection; hereditary;  
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;  
KW treatment; marker; ss.  
OS Homo sapiens.  
PN WO9820165-A2.  
PD 14-MAY-1998.  
PF 05-NOV-1997; U20313.  
PR 06-NOV-1996; US-030455.  
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
PI Hudson T, Lander ES, Wang D;  
DR WPI; 98-286974/25.  
PT New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease  
PS Claim 1; Page 219; 310pp; English.  
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
CC markers which have been isolated using the primers represented in  
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.  
SQ Sequence 100 BP; 21 A; 25 C; 22 G; 31 T;

Query Match 0.2%; Score 64; DB 1; Length 100;  
Best Local Similarity 77.6%; Pred. No. 0.12;  
Matches 76; Conservative 1; Mismatches 21; Indels 0; Gaps 0;

QY 4303 GTGGCTCAGGCTGTAAATCCCAAGCACTTCAGGAGGCTGAGTTGGGAGATCCTTGAGCT 4362  
DB 99 GTGACTCACACCTATATCTCTGGCACTTTAGGAGGCTTAGGAGGAGGATGTTTGAAC 40

QY 4363 CAGGAGTTCAAGACCACTTTTGGGCAACATAGCAAGTCT 4400  
DB 99 GTGACTCACACCTATATCTCTGGCACTTTAGGAGGCTTAGGAGGAGGATGTTTGAAC 40

DB 39 CAGGAGCTCAAGACCATCTCTGGGAAACATAGCAAGACT 2

RESULT 6  
X12087/c  
ID X12087 standard; DNA; 100 BP.  
AC X12087;  
DE 30-MAR-1999 (first entry)  
KW Human biallelic polymorphic DNA fragment EST98276a.  
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;  
KW detection; phenotypic typing; characteristic; infection; hereditary;  
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;  
KW treatment; marker; ss.  
OS Homo sapiens.  
PN WO9820165-A2.  
PD 14-MAY-1998.  
PF 05-NOV-1997; U20313.  
PR 06-NOV-1996; US-030455.  
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
PI Hudson T, Lander ES, Wang D;  
DR WPI; 98-286974/25.  
PT New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease  
PS Claim 1; Page 219; 310pp; English.  
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
CC markers which have been isolated using the primers represented in  
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.  
SQ Sequence 100 BP; 21 A; 25 C; 22 G; 31 T;

Query Match 0.2%; Score 63.6; DB 1; Length 100;  
Best Local Similarity 78.1%; Pred. No. 0.14;  
Matches 75; Conservative 1; Mismatches 20; Indels 0; Gaps 0;

QY 22822 GAGGCTCTGCTGTAAATCCCAAGCACTTTGGGAGGCGGAGGAGGATGCTTGAGCC 22881  
DB 99 GTGACTCACACCTATATCTCTGGCACTTTAGGAGGCTTAGGAGGAGGATGTTTGAAC 40

QY 22882 CAGGAATCAACACCAGCTCTGGGAAACATAGGAGAG 22917  
DB 39 CAGGAGCTCAAGACCAKCCCTGGGAAACATAGCAAGA 4

RESULT 7  
X12085/c  
ID X12085 standard; DNA; 100 BP.  
AC X12085;  
DE 30-MAR-1999 (first entry)  
KW Human biallelic polymorphic DNA fragment EST98276c.  
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;  
KW detection; phenotypic typing; characteristic; infection; hereditary;  
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;  
KW treatment; marker; ss.  
OS Homo sapiens.  
PN WO9820165-A2.  
PD 14-MAY-1998.





1000



PI Matsubara K, Okubo K;  
DR WPI; 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1; Page 1748; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 108 BP; 34 A; 31 C; 26 G; 15 T;

Query Match 0.2%; Score 57.8; DB 1; Length 108;  
Best Local Similarity 72.5%; Pred. No. 1;  
Matches 74; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 7242 ATTGCTTGAAACCCAGGAGGAGGTTGCAGTGAGCCGAGATCGCCACTGCATTCCAG 7301  
DB 1  
2 ATGCGCTGAGCCCATGAGGCCAAGGTCGAGTGCGCATGTCACGCCACTGNATCCAG 61

QY 7302 CCTGGCTGATAGTGCAGTCCATCTCGAGAAAAA 7343  
DB 1  
62 CCTGAGTGAGAGCAAGACCCCTGTTGAAAAACAACAACA 103



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2000, 15:22:36 ; Search time 8512.35 seconds  
(without alignments)  
13809.057 Million cell updates/sec

Title: US-08-852-495C-1\_COPY\_28000\_57000  
Perfect score: 29001  
Sequence: 1 ATATCACACAAAACACACAT.....TTAGCAGCACACAGGTAGGGT 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues  
Total number of hits satisfying chosen parameters: 156056

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : EST:  
1: em\_est1:  
2: em\_est2:  
3: em\_est3:  
4: em\_est4:  
5: em\_est5:  
6: em\_est6:  
7: em\_est7:  
8: em\_est8:  
9: em\_est9:  
10: em\_est10:  
11: em\_est11:  
12: em\_est12:  
13: em\_est13:  
14: em\_est14:  
15: em\_est15:  
16: em\_est16:  
17: em\_est17:  
18: em\_est18:  
19: em\_est19:  
20: gb\_est1:  
21: gb\_est2:  
22: gb\_est3:  
23: gb\_est4:  
24: gb\_est5:  
25: gb\_est6:  
26: gb\_est7:  
27: gb\_est8:  
28: gb\_est9:  
29: gb\_est10:  
30: gb\_est11:  
31: gb\_est12:  
32: gb\_est13:  
33: gb\_est14:  
34: gb\_est15:  
35: gb\_est16:  
36: gb\_est17:  
37: gb\_est18:  
38: gb\_est19:  
39: gb\_est20:  
40: gb\_est21:  
41: gb\_est22:  
42: gb\_est23:  
43: gb\_est24:  
44: gb\_est25:

45: gb\_est26:  
46: gb\_est27:  
47: gb\_est28:  
48: gb\_est29:  
49: gb\_est30:  
50: gb\_est31:  
51: gb\_est32:  
52: em\_est20:  
53: em\_est21:  
54: em\_est22:  
55: em\_est23:  
56: em\_est24:  
57: em\_est25:  
58: em\_est26:  
59: gb\_est33:  
60: gb\_est34:  
61: gb\_est35:  
62: gb\_est36:  
63: gb\_est37:  
64: gb\_est38:  
65: em\_est27:  
66: em\_est28:  
67: em\_est29:  
68: em\_est30:  
69: gb\_est39:  
70: gb\_est40:  
71: gb\_est41:  
72: gb\_est42:  
73: gb\_est43:  
74: gb\_est44:  
75: em\_est31:  
76: em\_est32:  
77: em\_est33:  
78: em\_est34:  
79: gb\_est45:  
80: gb\_est46:  
81: gb\_est47:  
82: gb\_gss1:  
83: gb\_gss2:  
84: gb\_gss3:  
85: gb\_gss4:  
86: em\_gss1:  
87: em\_gss2:  
88: em\_gss3:  
89: em\_gss4:  
90: gb\_gss5:  
91: gb\_gss6:  
92: gb\_gss7:  
93: gb\_gss8:  
94: gb\_gss9:  
95: em\_gss5:  
96: em\_gss6:  
97: em\_gss7:  
98: em\_gss8:  
99: em\_gss9:  
100: em\_gss10:  
101: em\_gss11:  
102: gb\_gss10:  
103: gb\_gss11:  
104: em\_gss12:  
105: gb\_gss12:  
106: gb\_gss13:  
107: gb\_gss14:  
108: gb\_gss15:  
109: gb\_gss16:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query



## FEATURES

source Location/Qualifiers

1. .109  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="A-345K02"  
 /clone\_lib="CIT978SKA1"  
 /sex="Female"  
 /cell\_type="Fibroblast"  
 /note="vector: pBAC108L; Site\_1: HindIII; Site\_2: HindIII;  
 CalTech Human BAC Library A1"  
 24 a 30 c 31 g 24 t

## BASE COUNT

24 a 30 c 31 g 24 t

Query Match 0.3%; Score 93; DB 84; Length 109;  
 Best Local Similarity 90.8%; Pred. No. 0.22; Indels 0; Gaps 0;  
 Matches 99; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 21860 TAGTAGACGGGGTTTCACCATGTTGCTCAGGCTGGTCTGGAACCTCTGACCTCAGGTG 21919

Db 109 TAGTTGACGGGGTTTCACCATGTTGCTCAGGCTGGTCTCGAAGCTCCGACCTCAGGTG 50

Qy 21920 ATTCGCCCACTCAGCTCCCAAGTCTGGGATTACAGGATGAGCCA 21968

Db 49 ATCCGCCACATCAGCTCCCAAGTCTGGGATTATAGGTATGAGCCA 1

## RESULT 3

## A1832832

LOCUS A1832832 105 bp mRNA EST 13-JUL-1999  
 DEFINITION at72909.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone  
 IMAGE:2377600 3' similar to contains Alu repetitive  
 element; contains element MER22 repetitive element ;, mRNA sequence.  
 A1832832  
 ACCESSION A1832832.1 GI:5454812  
 VERSION EST.  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 105)  
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
 Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
 Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
 WashU-NCI human EST Project  
 Unpublished (1997)  
 On Dec 20, 1995 this sequence version replaced gi:1133644.  
 Contact: Wilton RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40UP from Gibco.

## FEATURES

source Location/Qualifiers

1. .105  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2377600"  
 /clone\_lib="Barstead colon HPLRB7"  
 /sex="male"  
 /dev\_stage="adult, age 25"  
 /lab\_host="DH10B (phage resistant)"  
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a  
 modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTACGAATCTGAAGTGGAGCGGCCCTTTTTTTTTTTTTTTTTTTTTTTT  
 3']; double stranded cDNA was ligated to Eco RI adaptors  
 [5' AATTCTACTAGTAA 3' and 5' ATTACTAGT 3'], digested  
 with Not I and cloned into the Not I and Eco RI sites of

the modified pT73 vector. Library constructed by Bob  
 Barstead." 17 a 35 c 27 g 26 t

Query Match 0.3%; Score 92.2; DB 61; Length 105;  
 Best Local Similarity 92.4%; Pred. No. 0.27;  
 Matches 97; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 5411 GAGATGGAGTTTCGCTCTTGTGCCCCAGGCTGGAGTCAATGGCGGATCTCGGCTCACC 5470  
 Db 1 GAGACAGAGTTTCGCTCTTGTGCCCCAGGCTGGAGTCAATGGCGGATCTCGGCTCACC 60

Qy 5471 GCACCTCTACCTCCAGGTTCAACGAATTCCTCCGCTCAGCCT 5515

Db 61 GCACCTCCACCTCCCGGTTCAACGAGATTCTCTGCTCAGCCT 105

## RESULT 4

## AA703592

LOCUS AA703592 106 bp mRNA EST 24-DEC-1997  
 DEFINITION ag81a10.r1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone  
 IMAGE:1140858 5' similar to contains Alu repetitive element; , mRNA  
 sequence.

ACCESSION AA703592.1 GI:2713610

VERSION EST.

KEYWORDS EST.

SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 106)  
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
 Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
 Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
 WashU-NCI human EST Project  
 Unpublished (1997)

## TITLE

## JOURNAL

## COMMENT

On Sep 12, 1996 this sequence version replaced gi:11397630.  
 Contact: Wilton RK

Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28ml3 revl ET from Amersham

High quality sequence stop: 53.

## FEATURES

## source

Location/Qualifiers  
 1. .106  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1140858"  
 /clone\_lib="Stratagene hNT neuron (#937233)"  
 /dev\_stage="hNT neurons"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; Cloned unidirectionally. Primer: Oligo dT.  
 Differentiated, post mitotic hNT neurons. Average insert  
 size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'  
 GAATTCGGCAGGAG 3' -3' adaptor sequence: 5'  
 CTCGAGTTTITTTTTTTTTTTTTTTT 3'.

19 a 29 c 29 g 29 t

## BASE COUNT

## ORIGIN

Query Match 0.3%; Score 91.6; DB 37; Length 106;  
 Best Local Similarity 91.5%; Pred. No. 0.32;  
 Matches 97; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 10900 TTTTATTAGATGGGGTTTCACCATGTTACCGAGGATGGTCTCGATCTCTGACCTCG 10959

Db	1	TTTTTAGTAGACGAGGTTTACCGGTGTAGCCAGGATGCTTCGATCTCCCTGCCTCG 60
QY	10960	TGATCCACCGCGTGTGGCGTCCCAAAGTCTCGGATTACAGGCGTG 11005
Db	61	TGATGCGCCCGCTCAGCTCCCAAGTCTCGGATTACAGGCGTG 106
RESULT	5	
AQ264176/c		
LOCUS		AQ264176 106 bp DNA GSS 27-OCT-1998
DEFINITION		CITBI-EL-2509A2.TF CITBI-El Homo sapiens genomic clone 2509A2, genomic survey sequence.
ACCESSION		AQ264176
VERSION		AQ264176.1 GI:3792743
KEYWORDS		GSS.
SOURCE		human.
ORGANISM		Homo sapiens
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS		1 (bases 1 to 106) Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.
TITLE		Use of a random human BAC End Sequence Database for Sequence-Ready Map Building
JOURNAL		Unpublished (1998)
COMMENT		Other GSSs: CITBI-El-2509A2.TR Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: mdamads@tigr.org Clones are available from Research Genetics (info@resgen.com). BAC clones and search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html. Seq primer: M13-21 Class: BAC ends.
FEATURES		Location/Qualifiers
source		1..106 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="2509A2" /clone_lib="CITBI-El" /sex="male" /cell_type="sperm" /note="vector: pBelOAC11; Site_1: EcoRI; Site_2: EcoRI; Caltech Human BAC Library D"
BASE COUNT		25 a 30 c 34 g 17 t
ORIGIN		
Query Match		0.3%; Score 91.6; DB 105; Length 106;
Best Local Similarity		91.5%; Pred. No. 0.32;
Matches		97; Conservative 0; Mismatches 9; Indels 0; Gaps 0
QY	10915	GGTTTTCACCATGTTAGCCAGGATGGTCTCGATCTCTGACCTGATCCACCCGCTTT 10974
Db	106	GGGTTTTCACCATGTTAGCCAGGACGCGCTTGTGCTCTGCTGATCCACCCGCTC 47
QY	10975	GGCTCCCAAAGTCTGGGATTACAGCGTGAGCCACCGTGCCCGG 11020
Db	46	GGTCTCCCAAAGTCTGGGATTACAGCGTGAGACTCTGCGCCCGG 1
RESULT	6	
AQ028426		
LOCUS		AQ028426 109 bp DNA GSS 30-JUN-1998
DEFINITION		CIT-HSP-2313G15.TF CIT-HSP Homo sapiens genomic clone 2313G15, genomic survey sequence.
ACCESSION		AQ028426

VERSION	AQ028426.1	GI:3268648
KEYWORDS	GSS.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
AUTHORS	Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE	1 (bases 1 to 109)	
JOURNAL	Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,	
COMMENT	Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C. Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998) Unpublished (1998) Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: mdadams@tigr.org Clones are available from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html. Seq primer: M13-21 Class: BAC ends.	
FEATURES	Location/Qualifiers	
source	1..109	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="2313G15"	
	/clone_lib="citr-HSP"	
	/sex="Male"	
	/cell_type="Sperm"	
	/note="Vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII"	
BASE COUNT	19 a 36 c 25 g 29 t	
ORIGIN		
Query Match	0.3%; Score 91.4; DB 94; Length 109;	
Best Local Similarity	89.9%; Pred. No. 0.33; Indels 0; Gaps 0;	
Matches	90; Conservative 0; Mismatches 11;	
QY 15779	TTTTTTTTTGAGACAGAGTCTCACTCTATCACCCAGGCTGGAGTGCAGTGCACAAATCTC 15838	
Db 1	TTGTTTTCTGAGGGGACTCTCACTCTGTCAACCAGGCTGGAGTGCAGTGCACAGTCTG 60	
QY 15839	AGCTCACTGCAACCTGCACCTCTCGGGTTCAAGGGATTCTCCTACCTAA 15887	
Db 61	AGCTCACTGCAACCTGCACCTCTCGGGTTCAAGGGATTCTCCTGCTCTCA 109	
RESULT 7		
AQ535244	103 bp DNA GSS	18-MAY-1999
LOCUS	RPCI-11-31/H22.TV RPCI-11 Homo sapiens genomic clone	
DEFINITION	RPCI-11-31/H22, genomic survey sequence.	
ACCESSION	AQ535244	
VERSION	AQ535244.1	GI:4846934
KEYWORDS	GSS.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
	Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 103)	
AUTHORS	Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C. Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building Unpublished (1997) Contact: Shaying Zhao, William Nierman, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research	
TITLE		
JOURNAL		
COMMENT		

9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs (info@resgen.com). BAC end search page: [http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html).  
Seq primer: 77  
Class: BAC ends.

Location/Qualifiers

1. .103  
/organism="Homo sapiens"  
/db\_xref="GDB:7621533"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-317H22"  
/clone\_lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPCI11 Human Male BAC Library"

BASE COUNT  
ORIGIN

31 a 27 c 27 g 18 t  
Query Match 0.3%; Score 90.2; DB 108; Length 103;  
Best Local Similarity 92.2%; Pred. No. 0.45;  
Matches 95; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 7075 CCAGCACTTTGGGAGCGCAAAAGGCGGATCATTTGAGTTCAGGAGTTCGAGACCAAGCC 7134  
|||||  
Db 1 CCAGCACTTTGGGAGCGCAAGCGGAGATCATTTGAGTTCAGGAGTTCGAGACCAAGCC 60  
QY 7135 TGCCCAACATGGTGAACATCCATCTCTACTATAAATAACAAAAA 7177  
|||||  
Db 61 TGGCCAACATGGTGAACACCCGCTCTGCTATAAATAACAAAAA 103

RESULT 8  
AA565533/c

LOCUS AA565533 107 bp mRNA EST 08-SEP-1997  
DEFINITION nk42b11.s1 NCI\_CGAP\_GC2 Homo sapiens cDNA clone IMAGE:1016157 3'  
similar to contains Alu repetitive element;; mRNA sequence.  
AUTHORS AA565533  
ACCESSION AA565533  
VERSION AA565533.1 GI:2337172  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 107)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1393355.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Stratagene, Inc., David B. Kriman,  
Ph.D.  
CDNA Library Arraying: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/obrp/image/image.html](http://www-bio.llnl.gov/obrp/image/image.html)

Insert Length: 1661 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 87.  
Location/Qualifiers

source

1. .107  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1016157"  
/clone\_lib="NCI\_CGAP\_GC2"  
/tissue\_type="germ cell tumor"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: Bluescript SK; Site\_1: EcoRI; Site\_2:  
XhoI; cloned unidirectionally. Primer: Oligo dt. Bulk  
germ cell tumor. 5' adaptor sequence: 5' GAATTCGCACGAG 3'  
3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'  
Average insert size: 1.2 kb."

BASE COUNT  
ORIGIN

22 a 34 c 26 g 25 t  
Query Match 0.3%; Score 89.4; DB 35; Length 107;  
Best Local Similarity 89.7%; Pred. No. 0.54;  
Matches 96; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 6265 TGGTGTGCTGTAGTCCAGCTACTCAGAGCTGGGCGAGGAGTAATCGCTTGAACCT 6324  
|||||  
Db 107 TGGTGTGCTGTATCCAGCTACTCAGAGGCTGAGGCACGAGTAATCATCTGAACCT 48  
QY 6325 GGGAGCGGAGATTGCAGTCCAGATCGCACCACCGCACTCCAG 6371  
|||||  
Db 47 GGGAGCGAGAGCTTGCAGTCCAGTTCAGATTCAGCCACTGCACCTCCAG 1

RESULT 9  
AA078003/c

LOCUS AA078003 105 bp mRNA EST 24-SEP-1999  
DEFINITION 7H12D08 Chromosome 7 HeLa cDNA Library Homo sapiens cDNA clone  
7H12D08, mRNA sequence.  
ACCESSION AA078003  
VERSION AA078003.1 GI:1837477  
KEYWORDS EST.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 105)  
AUTHORS Touchman, J.W., Bouffard, G.G., Weintraub, L.A., Idol, J.R., Wang, L.,  
Robbins, C.M., Nussbaum, J.C., Lovett, M. and Green, E.D.  
TITLE 2006 expressed-sequence tags derived from human chromosome  
7-enriched cDNA libraries  
JOURNAL Genome Res. 7 (3), 281-292 (1997)  
MEDLINE 97228905  
COMMENT On Apr 14, 1993 this sequence version replaced gi:693433.  
Contact: Eric D. Green  
Genome Technology Branch  
National Human Genome Research Institute/NIH  
49 Convent Dr., MSC4431, Building 49, Room 2A08, Bethesda, MD 20892  
Tel: 3014020201  
Fax: 3014024735  
Email: egreen@nhgri.nih.gov  
Plate: 12 row: D column: 08  
Seq primer: -21M13 (ABI).

FEATURES  
source

1. .105  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="7H12D08"  
/clone="Chromosome 7 HeLa cDNA Library"  
/sex="female"  
/cell\_line="HeLa cell line: ATCC"  
/lab\_host="E. coli strain DH5 alpha"  
/note="Vector: pAMP10; cDNA was generated from cytoplasmic  
RNA using a mixture of random DNA hexamers and oligo(dT).  
From this pool of cDNA, human chromosome 7-enriched cDNA  
was isolated by direct cDNA selection using chromosome 7

genomic DNA (cosmids). The resulting direct-selected cDNA was cloned into a plasmid vector using a non-directional uracil DNA glycosylase (UDG)-mediated cloning strategy."

## BASE COUNT

20 a

33 c 23 g

29 t

## Query Match

0.3%; Score 89; DB 28; Length 105;

Best Local Similarity 90.5%; Pred. No. 0.61;

Matches 95; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

## QY

9922 TACAAAAATTAGCGGGCATGGTGGTGCACGCCCTCTAATCCAGCTACTTGGGAGGCTGA 9981

## Db

105 TACAAAAATTAGCTGGCATGTAGCGACGCATGTAATCCAGCTACTTGGGAGGCTGA 46

## QY

9982 GACAGAGATCGCTTGACCTGGGAGGAGGAGTTACAGTGAGC 10026

## Db

45 GACCGAGAAATGCTTGAACCCAGGAGGAGGTTGCAGTGAGC 1

## RESULT 10

## AI832832/c

## LOCUS

AI832832 105 bp mRNA EST 13-JUL-1999

## DEFINITION

at72g09.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone

## IMAGE:2377600 3'

similar to contains Alu repetitive

## element; contains

element MER22 repetitive element ; mRNA sequence.

## ACCESSION

AI832832

## VERSION

AI832832.1 GI:5454812

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens

## Eukaryota;

Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

## Eutheria;

Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 105)

## AUTHORS

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,

Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,

Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.

## TITLE

WashU-NCI human EST Project

## JOURNAL

Unpublished (1997)

## COMMENT

On Dec 20, 1995 this sequence version replaced gi:1133644.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40UP from Gibco.

## FEATURES

Location/Qualifiers

1..105

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2377600"

/clone\_lib="Barstead colon HPLRB7"

/sex="male"

/dev\_stage="adult, age 25"

/lab\_host="DH10B (phage resistant)"

/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a

modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'

TGTTACGAATCTGAAGTGGGAGCGCGCCCTTTTCTTTTCTTTTCTTTTCTTTT

3']; double-stranded cDNA was ligated to Eco RI adaptors

[5', AATCTACTAGTAT 3' and 5' ATTACTAGT 3'], digested

with Not I and cloned into the Not I and Eco RI sites of

the modified pT73 vector. Library constructed by Bob

Barstead."

17 a

35 c 27 g 26 t

## Query Match

0.3%; Score 89; DB 61; Length 105;

Best Local Similarity 90.5%; Pred. No. 0.61;

Matches 95; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

## QY

20776 AAGCTGACACAGGAAATCGCTTGAACCTGGGAGCGGAGGTTGTGTCGAGCCGAGATCA 20835

## Db

105 AGGCTGAGGCGAGGAGAAATCGCTTGAACCCGGGAGGTTCGGTTCGAGCAAGATCG 46

## QY

20836 TGCCATTGCATCTCCAGCTGGGCAACAAGAGCGAAATCCGTCCTC 20880

## Db

45 CACCATTCATCTCCAGCTGGGCAACAAGAGCGAAATCTGTCTC 1

## RESULT 11

## B48914/c

## LOCUS

B48914 103 bp DNA GSS 08-APR-1999

## DEFINITION

RPC111-4A12.TP RPCI-11 Homo sapiens genomic clone RPCI-11-4A12,

genomic survey sequence.

## ACCESSION

B48914

## VERSION

B48914.1 GI:2601151

## KEYWORDS

GSS.

## SOURCE

human.

## ORGANISM

Homo sapiens

Eukaryota;

Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria;

Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 103)

## AUTHORS

Adams, M.D., Rounsley, S.D., Field, C.E., Bass, S., Linher, K.,

Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and

Venter, J.C.

## TITLE

Use of BAC End Sequences for Sequence-Ready Map Building

## JOURNAL

Unpublished (1997)

## COMMENT

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from

Research Genetics (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/hungen/bac\_end\_search/bac\_end\_search.html

Seq primer: SP6

Class: BAC ends.

## FEATURES

Location/Qualifiers

1..103

/organism="Homo sapiens"

/db\_xref="GDB:7501163"

/db\_xref="taxon:9606"

/clone="RPCI-11-4A12"

/clone\_lib="RPCI-11"

/sex="Male"

/cell\_type="Lymphocytes"

/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;

RPC111 Human Male BAC Library"

30 a

28 c

30 g

15 t

## BASE COUNT

ORIGIN

Query Match 0.3%; Score 88.6; DB 84; Length 103;

Best Local Similarity 91.3%; Pred. No. 0.68;

Matches 94; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

## QY

10896 TCTATTTTATTAGATGGGGTTTCACCATGTTAGCAGGATGTCCTCGATCTCTGAC 10955

## Db

103 TGTATTTTATGATAGACGGGGTTTCACCGTTTATAGCCGGGATGTCCTCGATCTCTGAC 44

## QY

10956 CTCGTGATCCACCGCTTTGGCTCCCAAGTGCTGGGATTAC 10998

## Db

43 CTCGTGATCCGCGCCCTCGGCTCCCAAGTGCTGGGCTTAC 1



```

RESULT 12
A0535244/c  A0535244      103 bp  DNA      GSS      18-MAY-1999
LOCUS      RCPI-11-317H22-TV RCPI-11 Homo sapiens genomic clone
DEFINITION RCPI-11-317H22, genomic survey sequence.
ACCESSION  A0535244
VERSION    A0535244.1 GI:4846934
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 103)
AUTHORS   Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and
            Venter, J.C.
TITLE     Use of BAC End Sequences from Library RCPI-11 for Sequence-Ready
            Map Building
JOURNAL   Unpublished (1997)
COMMENT   Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbeetigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.bufo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.bufo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.

FEATURES             Location/Qualifiers
     source           1..103
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2510E2"
                     /clone_lib="RCPI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
     note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
     RCPI11 Human Male BAC Library"
BASE COUNT  31 a 27 c 27 g 18 t
ORIGIN

Query Match      0.3%; Score 88.6; DB 108; Length 103;
Best Local Similarity 91.3%; Pred. No. 0.68;
Matches 94; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 21848 TTTTGTATTTTATAGACAGCGGGTTTACCAGTGTGGTCAGGCTGGTGTGAACCTCC 21907
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 103 TTTTGTATTTATAGACAGACGGGTTTACCAGTGTGGCCAGGCTGGTCTCGAAGCTCC 44

Qy 21908 TGACCTCAGGTGATCTGCCACCTCAGCCTCCCAAGTGTGG 21950
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 43 TGACCTCAAGTGATCTGCCCGCTTGTGGCCTCCCAAGTGTGG 1

RESULT 13
A0265749/c  A0265749      109 bp  DNA      GSS      27-OCT-1998
LOCUS      CITBI-E1-2510E2-TR CITBI-E1 Homo sapiens genomic clone 2510E2,
DEFINITION genomic survey sequence.
ACCESSION  A0265749
VERSION    A0265749.1 GI:3791503
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  1 (bases 1 to 102)
AUTHORS   Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K.,
            Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H.,
            Simon, M. and Venter, J.C.
TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
JOURNAL   Unpublished (1998)
COMMENT   Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org

REFERENCE  1 (bases 1 to 109)
AUTHORS   Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K.,
            Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and
            Venter, J.C.
TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
JOURNAL   Unpublished (1998)
COMMENT   Other GSSs: CITBI-E1-2510E2.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.

FEATURES             Location/Qualifiers
     source           1..109
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2510E2"
                     /clone_lib="CITBI-E1"
                     /sex="male"
                     /cell_type="sperm"
     note="Vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;
     CalTech Human BAC Library D"
BASE COUNT  24 a 29 c 26 g 30 t
ORIGIN

Query Match      0.3%; Score 88.2; DB 105; Length 109;
Best Local Similarity 88.1%; Pred. No. 0.73;
Matches 96; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 9889 GCCAACATGGTGAACCCCTGCTCTACTAAAATAACAAAATATTAGCCGGCATGGTGGT 9948
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 109 GCCAGCATGGTGAACCCCTGCTCTACTAGAAAATAACAAAATATTAGCGAGCGCTGGCG 50

Qy 9949 CACGGCTGTATCCAGCTACTTGGAGCTGAGACAGGAGAAATCGCTT 9997
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 49 CATGCCTGTATCCAGCTACTTGGGAAGCTGAGCAGGAGCATCACTT 1

RESULT 14
A0004934
LOCUS      CIT-HSP-2292A10.TF CIT-HSP Homo sapiens genomic clone 2292A10,
DEFINITION genomic survey sequence.
ACCESSION  A0004934
VERSION    A0004934.1 GI:3082379
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 102)
AUTHORS   Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K.,
            Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H.,
            Simon, M. and Venter, J.C.
TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
JOURNAL   Unpublished (1998)
COMMENT   Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org

```

Clones are available from Research Genetics (info@resgen.com). BAC  
end' search page:  
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html  
Seq primer: M13-21;  
Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
1. .102  
/organism="Homo sapiens"  
/db\_xref="GDB:7151269"  
/db\_xref="taxon:9606"  
/clone="2292A10"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2:  
HindIII"

BASE COUNT 17 a 39 c 26 g 20 t

Query Match 0.3%; Score 87.6; DB 94; Length 102;  
Best Local Similarity 91.2%; Pred. No. 0.87; Indels 0; Gaps 0;  
Matches 93; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 5433 GCCCAGCTGGAGTGCATGGCGGATCTCGGCTCACCGCAACCTCTACCTCCCGAGGTTTC 5492  
|||||  
Db 1 GCCCAGCTGGAGTGCATGGCGGATCTCGGCTCACCGCAACCTCTACCTCCCGAGGTTTC 60

QY 5493 AAGCAATTCCTCTGGCTCAGCCTCCCGAGTAGCTGGGATTAC 5534  
|||||

Db 61 AAGCGACTCTCTGGCTTAGGCTCCCGAGTAGCTGGCATTAC 102

## RESULT 15

AA807640 103 bp mRNA EST 05-MAR-1998  
LOCUS nx08b05.s1 NCI\_CGAP\_GC3 Homo sapiens cDNA clone IMAGE:1255473 3'  
DEFINITION similar to contains Alu repetitive element;; mRNA sequence.  
ACCESSION AA807640  
VERSION AA807640.1 GI:2877108  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 103)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2151346.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 774 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 87.

FEATURES  
source

Location/Qualifiers  
1. .103  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1255473"  
/clone\_lib="NCI\_CGAP\_GC3"  
/tissue\_type="pooled germ cell tumors"

/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker: 1st strand cDNA was prepared from 3 pooled  
germ cell tumors, and was then primed with a Not I -  
oligo(dT) primer. Double-stranded cDNA was ligated to Eco  
RI adaptors (Pharmacia), digested with Not I and cloned  
into the Not I and Eco RI sites of the modified pT7T3  
vector. Library is not normalized. Library was  
constructed by Bento Soares and M. Fatima Bonaldo. "  
BASE COUNT 19 a 27 c 30 g 27 t  
ORIGIN

Query Match 0.3%; Score 87.6; DB 38; Length 103;  
Best Local Similarity 91.2%; Pred. No. 0.87;  
Matches 93; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 10905 ATTAGAGATGGGGTTTCACCATGTTAGCCAGGATGGTCTCGATCTCCTGACGCTGATC 10964  
|||||  
Db 2 AGTAGAGATGGGGTTTCACCGTGTAGCCAGGATGGTCTCGATCTCCTGACGCTGATC 61

QY 10965 CACCCGGCTTGGCCTCCCAAGTCTGGGATTACAGGCTGA 11006  
|||||

Db 62 CGCTCACCTCGGCCTCCCAAGTCTGGGATTACAGGCTGA 103

Search completed: June 14, 2000, 23:50:22  
Job time: 40389 sec





Wed Jun 21 14:43:56 2000

us-08-852-495c-1\_copy\_28000\_57000.rni

Page 1

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2000, 20:28:41 ; Search time 372.61 seconds  
(without alignments)  
10117.003 Million cell updates/sec

Title: US-08-852-495C-1\_COPY\_28000\_57000  
Perfect score: 29001  
Sequence: 1 ATATCAACAAAAACACACAT.....TTAGCAGCACAGGTAGGT 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 374504

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	80	0.3	105	4	US-08-481-658B-65
C 2	80	0.3	105	4	US-08-477-504A-65
C 3	80	0.3	105	4	US-08-486-756A-65
C 4	80	0.3	105	4	US-08-485-862B-65
C 5	80	0.3	105	5	US-08-787-739-65
C 6	79.4	0.3	105	4	US-08-481-658B-65
C 7	79.4	0.3	105	4	US-08-477-504A-65
C 8	79.4	0.3	105	4	US-08-486-756A-65
C 9	79.4	0.3	105	4	US-08-485-862B-65
C 10	79.4	0.3	105	5	US-08-787-739-65
C 11	65.4	0.2	84	3	US-08-454-557C-91
C 12	65.4	0.2	84	4	US-08-340-426D-91
C 13	65.4	0.2	84	4	US-08-450-673C-91
C 14	65.4	0.2	84	6	PCT-US95-17111A-91
C 15	62.8	0.2	84	3	US-08-454-557C-91
C 16	62.8	0.2	84	4	US-08-340-426D-91
C 17	62.8	0.2	84	4	US-08-450-673C-91
C 18	62.8	0.2	84	6	PCT-US95-17111A-91
C 19	60.8	0.2	85	3	US-08-454-557C-92
C 20	60.8	0.2	85	4	US-08-340-426D-92
C 21	60.8	0.2	85	4	US-08-450-673C-92
C 22	60.8	0.2	85	6	PCT-US95-17111A-92
C 23	60.4	0.2	78	3	US-08-454-557C-70
C 24	60.4	0.2	78	4	US-08-340-426D-70
C 25	60.4	0.2	78	4	US-08-450-673C-70
C 26	60.4	0.2	78	6	PCT-US95-17111A-70
C 27	58.8	0.2	78	3	US-08-454-557C-70

28	58.8	0.2	78	4	US-08-340-426D-70	Sequence 70, Appl
29	58.8	0.2	78	4	US-08-450-673C-70	Sequence 70, Appl
30	58.8	0.2	78	6	PCT-US95-17111A-70	Sequence 70, Appl
C 31	58	0.2	76	3	US-08-454-557C-69	Sequence 69, Appl
C 32	58	0.2	76	4	US-08-340-426D-69	Sequence 69, Appl
C 33	58	0.2	76	4	US-08-450-673C-69	Sequence 69, Appl
C 34	58	0.2	76	6	PCT-US95-17111A-69	Sequence 69, Appl
C 35	55.2	0.2	60	3	US-08-454-557C-57	Sequence 57, Appl
C 36	55.2	0.2	60	4	US-08-340-426D-57	Sequence 57, Appl
C 37	55.2	0.2	60	4	US-08-450-673C-57	Sequence 57, Appl
C 38	55.2	0.2	60	6	PCT-US95-17111A-57	Sequence 57, Appl
C 39	54.6	0.2	94	5	US-08-750-064-6	Sequence 6, Appl
40	53.6	0.2	60	3	US-08-454-557C-60	Sequence 60, Appl
41	53.6	0.2	60	4	US-08-340-426D-60	Sequence 60, Appl
42	53.6	0.2	60	4	US-08-450-673C-60	Sequence 60, Appl
43	53.6	0.2	60	6	PCT-US95-17111A-60	Sequence 60, Appl
44	51.6	0.2	76	3	US-08-454-557C-69	Sequence 69, Appl
45	51.6	0.2	76	4	US-08-340-426D-69	Sequence 69, Appl

## ALIGNMENTS

RESULT 1  
US-08-481-658B-65/C  
; Sequence 65, Application US/08481658B  
; Patent No 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,658B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.3E  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-481-658B-65

Query Match 0.3%; Score 80; DB 4; Length 105;  
Best Local Similarity 85.6%; Pred. No. 1.8e-09;

Matches 89; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 6147 ATCCGAGCAGCTTTGGAGGTGCGAGCGAGCTGATCAGGAGGTCAGAGGTTCAAGACCAGC 6206  
|||||  
Db 105 ATCCGAGCAGCTTTGGAGGTGCGAGCGAGGCTGGTGATCACAAGGTCAGAGGTTTGAGAGCAGC 46  
|||||

Qy 6207 CTGACCAAAATGATAAACCCCTGTCTCTACTAAAAATACAACA 6250  
|||||  
Db 45 CTGGCCAATATGGTGAACCCCTGTCTCTACTAAAAAGATGTAAAAA 2

## RESULT 2

US-08-477-504A-65/c  
; Sequence 65, Application US/08477504A  
; Patent No. 5972353  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/477,504A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.3D  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-477-504A-65

Query Match 0.3%; Score 80; DB 4; Length 105;  
Best Local Similarity 85.6%; Pred. No. 1.8e-09;  
Matches 89; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 6147 ATCCGAGCAGCTTTGGAGGTGCGAGCGAGCTGATCAGGAGGTCAGAGGTTCAAGACCAGC 6206  
|||||  
Db 105 ATCCGAGCAGCTTTGGAGGTGCGAGCGAGGCTGGTGATCACAAGGTCAGAGGTTTGAGAGCAGC 46  
|||||

Qy 6207 CTGACCAAAATGATAAACCCCTGTCTCTACTAAAAATACAACA 6250  
|||||  
Db 45 CTGGCCAATATGGTGAACCCCTGTCTCTACTAAAAAGATGTAAAAA 2

## RESULT 3

US-08-486-756A-65/c  
; Sequence 65, Application US/08486756A  
; Patent No. 5981711  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/486,756A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.3C  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-486-756A-65

Query Match 0.3%; Score 80; DB 4; Length 105;  
Best Local Similarity 85.6%; Pred. No. 1.8e-09;  
Matches 89; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 6147 ATCCGAGCAGCTTTGGAGGTGCGAGCGAGCTGATCAGGAGGTCAGAGGTTCAAGACCAGC 6206  
|||||  
Db 105 ATCCGAGCAGCTTTGGAGGTGCGAGCGAGGCTGGTGATCACAAGGTCAGAGGTTTGAGAGCAGC 46  
|||||

Qy 6207 CTGACCAAAATGATAAACCCCTGTCTCTACTAAAAATACAACA 6250  
|||||  
Db 45 CTGGCCAATATGGTGAACCCCTGTCTCTACTAAAAAGATGTAAAAA 2

## RESULT 4

US-08-485-862B-65/c  
; Sequence 65, Application US/08485862B  
; Patent No. 5989838  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court

CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,862B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-485-862B-65

Query Match 0.3%; Score 80; DB 4; Length 105;  
Best Local Similarity 85.6%; Pred. No. 1.8e-09;  
Matches 89; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 6147 ATCCGAGCACTTTGGAGGTCGAGCGCTGATCAGAGGTCAGAGTTCAAGACCAGC 6206  
Db 105 ATCCGAGCACTTTGGAGGTCGAGCGCTGATCAGAGGTCAGAGTTCAAGACCAGC 46  
QY 6207 CTGACCAAAATGATGAACCCCTGCTCTACTAAATAACAACA 6250  
Db 45 CTGGCCAATATGGTGAACCCCTGCTCTACTAAAGATGTAAAAA 2

RESULT 5  
US-08-787-739-65/C  
; Sequence 65, Application US/08787739  
; Patent No. 6027887  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 96  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 369 Pine Street, Suite 610  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/787,739  
; FILING DATE: 24-JAN-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,049  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/486,756  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/477,504  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/481,658  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,862  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,863  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/487,077  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-981-0332  
; TELEFAX: 415-981-0332  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; US-08-787-739-65

Query Match 0.3%; Score 80; DB 5; Length 105;  
Best Local Similarity 85.6%; Pred. No. 1.8e-09;  
Matches 89; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 6147 ATCCGAGCACTTTGGAGGTCGAGCGCTGATCAGAGGTCAGAGTTCAAGACCAGC 6206  
Db 105 ATCCGAGCACTTTGGAGGTCGAGCGCTGATCAGAGGTCAGAGTTCAAGACCAGC 46  
QY 6207 CTGACCAAAATGATGAACCCCTGCTCTACTAAATAACAACA 6250  
Db 45 CTGGCCAATATGGTGAACCCCTGCTCTACTAAAGATGTAAAAA 2

RESULT 6  
US-08-481-658B-65  
; Sequence 65, Application US/08481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:





MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-486-756A-65

Query Match 0.3%; Score 79.4; DB 4; Length 105;  
Best Local Similarity 84.8%; Pred. No. 2.5e-09;  
Matches 89; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 10891 TTTTGTATTTTATTAGAGATGGGGTTTCACCATGTTAGCCAGGATGGTCTCGATCTC 10950  
DB 1 TTTTGTACATCTTTAGTAGAGACAGGGTTTCACCATATTTGGCCAGGCTCTCTCAAACTC 60  
QY 10951 CTGACCTCGTGATCCACCGCTTTGGCCCTCCCAAAAGTGTCTGGGAT 10995  
DB 61 CTGACCTTGATGATCCACCGCTCGGCCCTCCCAAAAGTGTCTGGGAT 105

RESULT 9  
US-08-485-862B-65  
; Sequence 65, Application US/08485862B  
; Patent No. 5989838  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485.862B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/477,504  
; FILING DATE: 07-JUN-1995  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.3D  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-485-862B-65

Query Match 0.3%; Score 79.4; DB 4; Length 105;  
Best Local Similarity 84.8%; Pred. No. 2.5e-09;  
Matches 89; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 10891 TTTTGTATTTTATTAGAGATGGGGTTTCACCATGTTAGCCAGGATGGTCTCGATCTC 10950  
DB 1 TTTTGTACATCTTTAGTAGAGACAGGGTTTCACCATATTTGGCCAGGCTCTCTCAAACTC 60  
QY 10951 CTGACCTCGTGATCCACCGCTTTGGCCCTCCCAAAAGTGTCTGGGAT 10995  
DB 61 CTGACCTTGATGATCCACCGCTCGGCCCTCCCAAAAGTGTCTGGGAT 105  
RESULT 10  
US-08-787-739-65  
; Sequence 65, Application US/08787739  
; Patent No. 6027887  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 96  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 369 Pine Street, Suite 610  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/787,739  
; FILING DATE: 24-JAN-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,049  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/486,756  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/477,504  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/481,658  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,862  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,863  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/487,077  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-981-2034  
; TELEFAX: 415-981-0332  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-787-739-65

```
Query Match      0.3%; Score 79.4; DB 5; Length 105;
Best Local Similarity 84.8%; Pred. No. 2.5e-09;
Matches 89; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 10891 TTTTGTGATTTTATTAGAGAGGGTTTCACCATGTTAGCCAGGATGGTCTCGATCTC 10950
        ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 TTTTGTACATCTTAGTAGAGACAGGGTTTCACCATATTTGCCAGGCTGCTCTCAAACTC 60

Qy 10951 CTGACCTGTCATCCACCGCTTTGGCTCCCAAGTCTGGGAT 10995
        ||||| ||||| ||||| || ||||| ||||| ||||| ||||| |||||
Db 61 CTGACCTGTGTCACACAGCCTCGGCTCCCAAGTCTGGGAT 105

RESULT 11
US-08-454-557C-91
; Sequence 91, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 30-MAY-1995
; APPLICATION NUMBER: US/08/454,557C
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2540
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; US-08-454-557C-91

Query Match      0.2%; Score 65.4; DB 3; Length 84;
Best Local Similarity 86.7%; Pred. No. 3.5e-06;
Matches 72; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 10923 CCATGTTAGCCAGGATGGTCTCGATCTCCTGACCTCGTGATCCACCGCTTTGGCCTCCC 10982
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 CCATGTTTCATCAGGCTGGTGTGCGAATCCTGACCTCGTGATCCCGCCCTCAGCCTCCC 60

Qy 10983 AAAGTCTGGGATTACAGCGCTG 11005
        ||||| ||||| ||||| ||||| |||||
Db 61 AAAGTCTGGGATTACAGCGTG 83

RESULT 12
US-08-340-426D-91
; Sequence 91, Application US/08340426D
; Patent No. 5948634
```

```
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/340,426D
; FILING DATE: 14-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2540
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; US-08-340-426D-91

Query Match      0.2%; Score 65.4; DB 4; Length 84;
Best Local Similarity 86.7%; Pred. No. 3.5e-06;
Matches 72; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 10923 CCATGTTAGCCAGGATGGTCTCGATCTCCTGACCTCGTGATCCACCGCTTTGGCCTCCC 10982
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 CCATGTTTCATCAGGCTGGTGTGCGAATCCTGACCTCGTGATCCCGCCCTCAGCCTCCC 60

Qy 10983 AAAGTCTGGGATTACAGCGCTG 11005
        ||||| ||||| ||||| ||||| |||||
Db 61 AAAGTCTGGGATTACAGCGTG 83

RESULT 13
US-08-450-673C-91
; Sequence 91, Application US/08450673C
; Patent No. 5948888
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
```

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450.673C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
US-08-450-673C-91

Query Match 0.2%; Score 65.4; DB 4; Length 84;
Best Local Similarity 86.7%; Pred. No. 3.5e-06;
Matches 72; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 10923 CCATGTTAGCCAGGATGCTCGATCTCCTGACCTCGTGATCCACCGCTTTGGCCTCCC 10982
Db 1 CCATGTTTCATCAGGCTGCTCGAACTCCTGACCTCGTGATCCCGCGCTCAGCCTCCC 60

Qy 10983 AAAGTGCTGGGATTACAGCGGTG 11005
Db 61 AAAGTGCTGGGATTACAGCGGTG 83

RESULT 14
PCT-US95-17111A-91
; Sequence 91, Application PC/TUS9517111A
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and
; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
US-08-450-673C-91

Query Match 0.2%; Score 62.8; DB 3; Length 84;
Best Local Similarity 85.4%; Pred. No. 1.4e-05;
Matches 70; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 6137 CATGACTGTATATCCAGCACCTTTGGGAGGTCGAGGAGGCTGATCAGGAGTCAGAGTT 6196
Db 83 CACGCTTGTATATCCAGCACCTTTGGGAGGCTGAGGAGGTCAGGAGTCAGGAGTT 24

Qy 6197 CAAGACACGCTGACCAAAATG 6218
Db 23 CGACACACGCTGATGATCATG 2
```

```
;
; TOPOLOGY: both
PCT-US95-17111A-91

Query Match 0.2%; Score 65.4; DB 6; Length 84;
Best Local Similarity 86.7%; Pred. No. 3.5e-06;
Matches 72; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 10923 CCATGTTAGCCAGGATGCTCGATCTCCTGACCTCGTGATCCACCGCTTTGGCCTCCC 10982
Db 1 CCATGTTTCATCAGGCTGCTCGAACTCCTGACCTCGTGATCCCGCGCTCAGCCTCCC 60

Qy 10983 AAAGTGCTGGGATTACAGCGGTG 11005
Db 61 AAAGTGCTGGGATTACAGCGGTG 83

RESULT 15
US-08-454-557C-91/c
; Sequence 91, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,557C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
US-08-454-557C-91

Query Match 0.2%; Score 62.8; DB 3; Length 84;
Best Local Similarity 85.4%; Pred. No. 1.4e-05;
Matches 70; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 6137 CATGACTGTATATCCAGCACCTTTGGGAGGTCGAGGAGGCTGATCAGGAGTCAGAGTT 6196
Db 83 CACGCTTGTATATCCAGCACCTTTGGGAGGCTGAGGAGGTCAGGAGTCAGGAGTT 24

Qy 6197 CAAGACACGCTGACCAAAATG 6218
Db 23 CGACACACGCTGATGATCATG 2
```

Search completed: June 15, 2000, 04:56:24  
Job time: 57714 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2000, 04:50:04 ; Search time 17971.5 Seconds  
(without alignments)  
-1569.819 Million cell updates/sec

Title: US-08-852-495c-1\_copy\_56000\_85000  
Perfect score: 29001  
Sequence: 1 ATGACAAAGGCTAGTGAT.....CAGGAGACTAGAGTTTTATT 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 370290

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_bal.\*
- 2: gb\_ba2.\*
- 3: gb\_on.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pl1.\*
- 8: gb\_pl2.\*
- 9: gb\_pr1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_vl.\*
- 17: em\_fun.\*
- 18: em\_hum1.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_om.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pl.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_vl.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_bal.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*
- 44: gb\_htg6.\*

- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*
- 50: gb\_pl3.\*
- 51: gb\_pr5.\*
- 52: gb\_htg8.\*
- 53: gb\_htg9.\*
- 54: gb\_htg10.\*
- 55: gb\_htg11.\*
- 56: gb\_htg12.\*
- 57: gb\_htg13.\*
- 58: gb\_htg14.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	92	0.3	108	10	HSLDLRN2
2	92	0.3	108	10	HSLDLRN2
3	88.6	0.3	107	9	HUMALCE162
4	88	0.3	108	11	HSU67803
5	84	0.3	103	9	HUMALCE221
6	83.6	0.3	103	9	HUMALCE221
7	83.8	0.3	104	9	HUMALCE272
8	83	0.3	108	10	HSLDLRD1
9	83	0.3	108	10	HSLDLRD2
10	81.4	0.3	108	10	HSLDLRD1
11	81.4	0.3	108	10	HSLDLRD2
12	78.2	0.3	91	13	HUMUT8164A
13	78.4	0.3	108	10	HSLDLI12
14	77.8	0.3	108	11	HSU67803
15	77.8	0.3	110	9	HUMALCE43
16	76.2	0.3	108	11	HSU67808
17	76.4	0.3	108	11	HSU67808
18	76.4	0.3	110	9	HUMALCE43
19	76	0.3	104	9	HUMALCE272
20	75.8	0.3	107	9	HUMALCE162
21	75.4	0.3	106	13	G32743
22	75.2	0.3	108	11	HSU67804
23	74.4	0.3	103	13	HS8IC8R
24	73.4	0.3	103	13	HS8IC8R
25	72	0.2	90	9	HUMLDLRFL
26	72	0.2	108	10	HSLDLI12
27	71.4	0.2	107	11	HSU67806
28	71.4	0.2	108	11	HSU67804
29	70.8	0.2	91	13	HUMUT8164A
30	70	0.2	107	11	HSU67806
31	70	0.2	108	9	HUMD1D03M5
32	70.2	0.2	110	11	HSU67807
33	70.2	0.2	110	11	HSU67807
34	69.4	0.2	97	9	HUMLDLRA2
35	69.4	0.2	100	9	HUMGALNSA
36	68.8	0.2	85	9	HUMBRKFAE
37	68.4	0.2	90	13	HUMUT8002B
38	68.4	0.2	102	13	G32906
39	67.6	0.2	101	10	S79560
40	67.8	0.2	108	9	HUMD1D03M5
41	67.2	0.2	84	5	AR051521
42	67.2	0.2	90	9	HUMLDLRFL
43	66.8	0.2	102	13	G32906
44	66	0.2	100	10	HSLAS27
45	66	0.2	100	13	HUMUT931A

ALIGNMENTS

```

RESULT 1
HSLDLRN2
LOCUS      HSLDLRN2      108 bp      DNA      PRI      20-MAY-1992
DEFINITION Human LDL-receptor gene intron 14 fragment (normal gene).
ACCESSION X05250
VERSION    X05250.1 GI:34337
KEYWORDS   Alu repetitive sequence; low density lipoprotein receptor.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 108)
AUTHORS    Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
            Williamson,R. and Humphries,S.
TITLE      Unequal crossing-over between two alu-repetitive DNA sequences in
            the low-density-lipoprotein-receptor gene. A possible mechanism for
            the defect in a patient with familial hypercholesterolaemia
JOURNAL    Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE    87161901
COMMENT     See X05252 for deletion junction
            Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES   Location/Qualifiers
            source      1..108
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
            intron      1..108
                        /note="intron XIV fragment"
BASE COUNT 28 a 23 c 39 g 18 t
ORIGIN

Query Match      0.3%; Score 92; DB 10; Length 108;
Best Local Similarity 90.7%; Pred. No. 2.le-06;
Matches 98; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 10989 ACRAAGTAGTGGCGGTGGACATCGCTAGTCCAGCTACTGGGGAGGCTGAG 11048
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ACRAAATTAGCCAGCGGTGGTGGCAGGTGCTGTATCCAGCTACTCGGAGGCTGAG 60
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 11049 GCAGGAGAAATGCTGAACTCGGGAGCGGAGGTGTCAGTGCAGCCGAG 11096
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GCAGGAGAAATGCTTGAACCCAGGAGGACAGAGGTTCGAGTGAGCCGAG 108
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
HSLDLRN2/c
LOCUS      HSLDLRN2      108 bp      DNA      PRI      20-MAY-1992
DEFINITION Human LDL-receptor gene intron 14 fragment (normal gene).
ACCESSION X05250
VERSION    X05250.1 GI:34337
KEYWORDS   Alu repetitive sequence; low density lipoprotein receptor.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 108)
AUTHORS    Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
            Williamson,R. and Humphries,S.
TITLE      Unequal crossing-over between two alu-repetitive DNA sequences in
            the low-density-lipoprotein-receptor gene. A possible mechanism for
            the defect in a patient with familial hypercholesterolaemia
JOURNAL    Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE    87161901
COMMENT     See X05252 for deletion junction
            Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES   Location/Qualifiers
            source      1..108
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
            intron      1..108
                        /note="intron XIV fragment"
BASE COUNT 28 a 23 c 39 g 18 t
ORIGIN

```

## ORIGIN

```

Query Match      0.3%; Score 92; DB 10; Length 108;
Best Local Similarity 90.7%; Pred. No. 2.le-06;
Matches 98; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 24355 CTCGGCTCACTCAACCTCGGCTCAGCGGTTCAAGCGATTCTCCTGCCCTCCGCTCCG 24414
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 108 CTCGGCTCACTCAACCTCGGCTCAGCGGTTCAAGCGATTCTCCTGCCCTCCGCTCCG 49
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 24415 AGTAGCTGAGATTACAGGCGCTGCCACCATCGCGGCTAAATTTTGT 24462
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 48 AGTAGCTGGGATTACAGGCGCTGCCACCATCGCGCTAAATTTTGT 1

RESULT 3
HUMALCE162/c
LOCUS      HUMALCE162      107 bp      ss-RNA      PRI      15-APR-1994
DEFINITION Human carcinoma cell-derived Alu RNA transcript, clone CE162.
ACCESSION M87924
VERSION    M87924.1 GI:174871
KEYWORDS   Alu repeat.
SOURCE     Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 107)
AUTHORS    Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE      Alu RNA transcripts in human embryonal carcinoma cells. Model of
            post-transcriptional selection of master sequences
JOURNAL    J. Mol. Biol. (1992) In press
FEATURES   Location/Qualifiers
            source      1..107
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /cell_line="Ntera2D1"
                        /dev_stage="embryo"
                        /sex="male"
                        /tissue_type="carcinoma"
BASE COUNT 28 a 30 c 35 g 14 t
ORIGIN

Query Match      0.3%; Score 88.6; DB 9; Length 107;
Best Local Similarity 91.3%; Pred. No. 8e-06;
Matches 94; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 11447 TGTTTGAGACGAGGAGTCTTCTGTGTCGCCAGGCTGAGTGCAGTGGTGTGATCTCCG 11506
        ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 107 TTTTGTGAGACGAGGAGTCTGCTGTGTCGCCAGGCTGAGTGCAGTGGCGGATCTCGC 48
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 11507 TCAGTGAAGCTCGGCTCCCGGATTACGCCATTCCTCTGCC 11549
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 47 TCAGTGAAGCTCGGCTCCCGGTTTCACGCCATTCCTCTGCC 5

RESULT 4
HSU67803/c
LOCUS      HSU67803      108 bp      RNA      PRI      01-AUG-1997
DEFINITION Human small cytoplasmic Alu transcript.
ACCESSION U67803
VERSION    U67803.1 GI:2289917
KEYWORDS   Alu.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 108)
AUTHORS    Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE      cDNAs derived from primary and small cytoplasmic Alu (scAlu)
            transcripts
JOURNAL    J. Mol. Biol. 271 (2), 222-234 (1997)

```

```

MEDLINE          97415756
REFERENCE        2 (bases 1 to 108)
AUTHORS          Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE            Direct Submission
JOURNAL          Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The
                  Children's Hospital of Philadelphia, 1004F Abramson Research
                  Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA
FEATURES
source
1..108
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TscAlu2"
1..108
/note="scAlu"
/rpt_family="Alu"
/rpt_type="dispersed"
BASE COUNT      23 a 39 c 30 g 16 t
ORIGIN

Query Match      0.3%; Score 88; DB 11; Length 108;
Best Local Similarity 94.8%; Pred. No. 1e-05;
Matches 91; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 11616 GTAGATACGGGGTTTCACTTTGTTAAACAGGATGGTCTCGATCTCTGACCTCGTGATCG 11675
Db 97 GTAGAGACGGGGTTTCACTTTGTTAGCCAGGATGGTCTCGATCTCTGACCTCGTGATCC 38

QY 11676 GCCCGCTCAGCTCCCAAGTGTGGGATTACAGG 11711
Db 37 GCCCGCTCGGCTCCCAAGTGTGGGATTACAGG 2

RESULT 5
HUMALCE221
LOCUS            HUMALCE221 103 bp ss-RNA PRI 15-APR-1994
DEFINITION      Human carcinoma cell-derived Alu RNA transcript, clone CE221.
ACCESSION       M87896
VERSION         M87896.1 GI:174874
KEYWORDS        Alu repeat.
SOURCE          Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                  Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1 (bases 1 to 103)
AUTHORS         Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE           Alu RNA transcripts in human embryonal carcinoma cells. Model of
                  post-transcriptional selection of master sequences
JOURNAL         J. Mol. Biol. (1992) In press
FEATURES
source
1..103
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Ntera2D1"
/dev_stage="embryo"
/sex="male"
/tissue_type="carcinoma"
BASE COUNT      25 a 27 c 33 g 18 t
ORIGIN

Query Match      0.3%; Score 84; DB 9; Length 103;
Best Local Similarity 90.0%; Pred. No. 4.9e-05;
Matches 90; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 11017 TGCCTGTAGTCCAGCTACTGGGGAGGCTGAGCAGAGAGATGCTTGAACTCGGGAGGC 11076
Db 4 TGCCTGTATCCAGCTACACGGGAGCTAAGCAGGAGATCGCTTGAACCGGGAGGC 63

QY 11077 GGAGGTTCAGTGCAGCCGAGATGCCCACTGCATCCAG 11116
Db 64 GGAGGTTCAGTGCAGCCGAGATCGTGCATTCGACTCCAG 103

```

```

RESULT 6
HUMALCE221/c
LOCUS            HUMALCE221 103 bp ss-RNA PRI 15-APR-1994
DEFINITION      Human carcinoma cell-derived Alu RNA transcript, clone CE221.
ACCESSION       M87896
VERSION         M87896.1 GI:174874
KEYWORDS        Alu repeat.
SOURCE          Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                  Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1 (bases 1 to 103)
AUTHORS         Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE           Alu RNA transcripts in human embryonal carcinoma cells. Model of
                  post-transcriptional selection of master sequences
JOURNAL         J. Mol. Biol. (1992) In press
FEATURES
source
1..103
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Ntera2D1"
/dev_stage="embryo"
/sex="male"
/tissue_type="carcinoma"
BASE COUNT      25 a 27 c 33 g 18 t
ORIGIN

Query Match      0.3%; Score 83.6; DB 9; Length 103;
Best Local Similarity 90.8%; Pred. No. 5.8e-05;
Matches 89; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 24335 CTGGAGTCTGTGGCAGCTTCTGGCTCACTGCAACCTCCGCTCAGCGTTCAAGCGAT 24394
Db 103 CTGGAGTGCATATGGCAGCATCTGGCTCACTGCAACCTCCGCTCAGCGTTCAAGCGAT 44

QY 24395 TCTCTGCTCCCTCGCCCTCCGAGTAGCTAGATTACAGG 24432
Db 43 TCTCTGCTCCCTAGCTTCCGCTGCTAGCTGGGATTACAGG 6

RESULT 7
HUMALCE272/c
LOCUS            HUMALCE272 104 bp ss-RNA PRI 15-APR-1994
DEFINITION      Human carcinoma cell-derived Alu RNA transcript, clone CE272.
ACCESSION       M87895
VERSION         M87895.1 GI:174875
KEYWORDS        Alu repeat.
SOURCE          Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                  Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1 (bases 1 to 104)
AUTHORS         Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE           Alu RNA transcripts in human embryonal carcinoma cells. Model of
                  post-transcriptional selection of master sequences
JOURNAL         J. Mol. Biol. (1992) In press
FEATURES
source
1..104
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Ntera2D1"
/dev_stage="embryo"
/sex="male"
/tissue_type="carcinoma"
BASE COUNT      22 a 26 c 37 g 19 t
ORIGIN

Query Match      0.3%; Score 83.8; DB 9; Length 104;
Best Local Similarity 88.3%; Pred. No. 5.3e-05;
Matches 91; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```





```

Query Match          0.3%; Score 81.4; DB 10; Length 108;
Best Local Similarity 85.0%; Pred. No. 0.00014;
Matches 91; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 10989 ACAAAGTTAGTCGGCGTGGTGACATGCTGTAGTCCAGCTACTGGGAGGCTGAG 11048
DB 108 ACAAATTTAGCCAGCGTGGTGGAGGTGCTGTAAATCCCACTACTCGGAGGCTGAG 49

QY 11049 GCAGGAGAAATGCTTGAACCTCGGAGCGGAGGTTGCAGTGAGCCGA 11095
DB 48 GCAGGAGAAATGCTTGAACCCAGGAGGAGAGGTTGTGTGTGAGGCCA 2

RESULT 11
HSLDLRD2          HSLDLRD2          108 bp DNA          PRI          20-MAY-1992
LOCUS             Human LDL-receptor mutated gene with intron 14 deletion junction.
DEFINITION
ACCESSION         X05251
VERSION           X05251.1 GI:34336
KEYWORDS          Alu repetitive sequence; low density lipoprotein receptor.
SOURCE            human.
ORGANISM           Homo sapiens
REFERENCE          Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS           Primates; Catarrhini; Hominidae; Homo.
TITLE             1 (bases 1 to 108)
                  Horsthemke, B., Beisiegel, U., Dunning, A., Hovinga, J.R.,
                  Williamson, R., and Humphries, S.
                  Unequal crossing-over between two alu-repetitive DNA sequences in
                  the low-density-lipoprotein-receptor gene. A possible mechanism for
                  the defect in a patient with familial hypercholesterolaemia
JOURNAL           Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE           87161901
COMMENT           *source: hypercholesterol aemia
                  See X05250 for corresponding normal gene sequence
                  In the defective LDL-receptor gene the deletion occurred between two
                  alu-repetitive sequences, that are in the same direction, the
                  deletion eliminates exons 13 and 14 and changes the reading frame
                  of the resulting spliced mRNA.
                  Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES           Location/Qualifiers
                   source
                     1..108
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="blood leukocytes from a patient with familial"
                   intron
                     1..108
                     /note="intron XIV fragment"
                   BASE COUNT 28 a 20 c 40 g 20 t
                   ORIGIN

Query Match          0.3%; Score 81.4; DB 10; Length 108;
Best Local Similarity 85.0%; Pred. No. 0.00014;
Matches 91; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 10989 ACAAAGTTAGTCGGCGTGGTGACATGCTGTAGTCCAGCTACTGGGAGGCTGAG 11048
DB 108 ACAAATTTAGCCAGCGTGGTGGAGGTGCTGTAAATCCCACTACTCGGAGGCTGAG 60

QY 11049 GCAGGAGAAATGCTTGAACCTCGGAGCGGAGGTTGCAGTGAGCCGA 11095
DB 61 GCAGGAGAAATGCTTGAACCCAGGAGGAGAGGTTGTGTGTGAGGCCA 107

RESULT 12
HUMUT8164A
LOCUS             HUMUT8164A          91 bp DNA          STS          29-DEC-1994
DEFINITION        Human STS UT8164, 5', primer bind, sequence tagged site.
ACCESSION         L30244
VERSION           L30244.1 GI:605447
KEYWORDS          STS; PCR primer; STS sequence; microsatellite DNA; microsatellite
SOURCE            marker; sequence tagged site; tetranucleotide repeat.
                   Homo sapiens DNA.

```

```

ORGANISM           Homo sapiens
REFERENCE          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS           Eutheria; Primates; Catarrhini; Hominidae; Homo.
                  1 (bases 1 to 91)
                  Gerken, S.C., Matsunami, N., Plaetke, R., Albertsen, H., Ballard, L.,
                  Mellis, R., Lawrence, E., Moore, M., Holik, P., Carlson, M., Zhao, X.,
                  Robertson, M., Bradley, P., Elsner, T., Tingey, A., Lalouel, J.-M., and
                  White, R.
                  Genetic and physical mapping of simple sequence repeat containing
                  sequence tagged sites from the human genome
JOURNAL           Unpublished (1994)
COMMENT           Submitted by: Utah Center for Human Genome Research University of
                  Utah, Dept. of Human Genetics
                  2160 Eccles Institute of Human Genetics
                  Salt Lake City, UT 84112
                  e-mail: sts@corona.med.utah.edu
                  Primer A: AGAGGTGCAGTGAAACCA
                  Primer B: TTTTCCCTCTACTCACT
                  End to Label: Primer B
                  PCR Profile:
                  Initial Denaturation: 94C 300sec
                  Cycles Denaturation Annealing Extension 5 94
                  C 10 sec. 56 C 10 sec. 72 C 20 sec. 30
                  52 C 10 sec. 72 C 20 sec. Mg++: 1.50 mM
                  Gel: Acrylamide 7%, Formamide 32%, Urea 34%
                  Alleles: 1.
FEATURES           Location/Qualifiers
                   source
                     1..91
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                   primer_bind
                     10..28
                   BASE COUNT 35 a 20 c 23 g 13 t
                   ORIGIN

Query Match          0.3%; Score 78.2; DB 13; Length 91;
Best Local Similarity 91.2%; Pred. No. 0.00049;
Matches 83; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 4131 TCAGGAGCAGAGGTTGCAGTGAGTGATCACACCATTCAGCTCCAGCTGGGTGACA 4190
DB 1 TCGGGAGGCAGAGGTTGCAGTGAGTGATCACACCATTCAGCTCCAGCTGGGTGACA 60

QY 4191 GTCGAGACTCTCTCAAAAAAATAAAAAA 4221
DB 61 GAGTGAGACTCTGTCAAAAAAATAAAAAA 91

RESULT 13
HSLDL112/c
LOCUS             HSLDL112          108 bp DNA          PRI          20-MAY-1992
DEFINITION        Human LDL-receptor gene intron 12 fragment (normal gene) LDL - low
                  density lipoprotein.
ACCESSION         X05248
VERSION           X05248.1 GI:34334
KEYWORDS          Alu repetitive sequence; low density lipoprotein receptor;
SOURCE            repetitive sequence.
                   human.
ORGANISM           Homo sapiens
REFERENCE          Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS           Primates; Catarrhini; Hominidae; Homo.
                  1 (bases 1 to 108)
                  Horsthemke, B., Beisiegel, U., Dunning, A., Hovinga, J.R.,
                  Williamson, R., and Humphries, S.
                  Unequal crossing-over between two alu-repetitive DNA sequences in
                  the low-density-lipoprotein-receptor gene. A possible mechanism for
                  the defect in a patient with familial hypercholesterolaemia
JOURNAL           Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE           87161901
COMMENT           see X05249 for deletion junction
                   Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES           Location/Qualifiers

```



---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2000, 05:38:04 ; Search time 593.79 seconds  
(without alignments)  
12219.506 Million cell updates/sec

Title: US-08-852-495C-1\_COPY\_56000\_85000  
Perfect score: 29001  
Sequence: 1 ATCAACAAGGCTGACTGAT.....CAGGAGACTAGACTTTTATT 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 433070

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	71	0.2	100	1 T24892	Human gene signatu
C 2	70.4	0.2	100	1 T24892	Human gene signatu
C 3	68.8	0.2	108	1 T26828	Human gene signatu
C 4	66.8	0.2	108	1 X12095	Human biallelic po
C 5	64.8	0.2	108	1 X12095	Human biallelic po
C 6	64.2	0.2	91	1 T25854	Human gene signatu
C 7	64.4	0.2	108	1 T26828	Human gene signatu
C 8	63.6	0.2	103	1 T26213	Human gene signatu
C 9	62.8	0.2	103	1 T20927	Human gene signatu
C 10	62.4	0.2	103	1 T20927	Human gene signatu
C 11	61	0.2	91	1 T25854	Human gene signatu
C 12	61	0.2	103	1 T26213	Human gene signatu
C 13	59.6	0.2	100	1 X12087	Human biallelic po
C 14	59.6	0.2	100	1 X12085	Human biallelic po
C 15	59.6	0.2	100	1 X12086	Human biallelic po
C 16	59	0.2	93	1 T22572	Human gene signatu
C 17	59.2	0.2	108	1 T25009	Human gene signatu
C 18	58.4	0.2	110	1 T26288	Human gene signatu
C 19	57.4	0.2	97	1 T26728	Human gene signatu
C 20	57.4	0.2	110	1 T26288	Human gene signatu
C 21	57	0.2	100	1 X12087	Human biallelic po
C 22	57	0.2	100	1 X12085	Human biallelic po
C 23	57	0.2	100	1 X12086	Human biallelic po
C 24	57	0.2	108	1 T25009	Human gene signatu
C 25	56	0.2	93	1 T25688	Human gene signatu
C 26	55.8	0.2	95	1 T23131	Human gene signatu
C 27	55	0.2	109	1 T23895	Human gene signatu
C 28	54.4	0.2	109	1 T23895	Human gene signatu
C 29	53.6	0.2	99	1 T20931	Human gene signatu
C 30	53.2	0.2	75	1 T22841	Human gene signatu
C 31	53.2	0.2	93	1 T22572	Human gene signatu
C 32	52.6	0.2	70	1 N60231	Normal chromosome
C 33	52.2	0.2	97	1 T26728	Human gene signatu
C 34	51.4	0.2	53	1 Q33621	Microsatellite seq

C 35	51.6	0.2	75	1 T22841	Human gene signatu
C 36	51.4	0.2	93	1 T24259	Human gene signatu
C 37	51	0.2	93	1 T24259	Human gene signatu
C 38	50.4	0.2	85	1 T26182	Human gene signatu
C 39	50.4	0.2	106	1 Q95210	Simple tandem repe
C 40	50	0.2	87	1 T21566	Human gene signatu
C 41	49.6	0.2	81	1 T24093	Human gene signatu
C 42	49.6	0.2	84	1 T25848	Human gene signatu
C 43	49.6	0.2	88	1 V39744	Microsatellite ana
C 44	49.6	0.2	93	1 T25688	Human gene signatu
C 45	49.8	0.2	100	1 T25604	Human gene signatu

ALIGNMENTS

RESULT 1  
T24892/C  
ID T24892 standard; CDNA to mRNA; 100 BP.  
AC T24892;  
DT 05-NOV-1996 (first entry)  
DE Human gene signature HUMGS06998.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN W09514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
WPI: 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1; Page 1720; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 100 BP; 28 A; 22 C; 25 G; 22 T;

Query Match 0.2%; Score 71; DB 1; Length 100;  
Best Local Similarity 81.6%; Pred. No. 0.015;  
Matches 80; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 17609 TTTTGTGTTTAAATAGAGTCTCGCTCTCTACCCAGCTGGAGTGCAGTGGCGCAATCT 17668

Db 99 TTGTTTGTGTTTAAATAGAGTCTCTCTCTACCCAGCTGGAGTGCAGTGGCGCAATCT 40

QY 17669 CAGCTCACTGCAACGTCGCCCTCTCGGGTTCAAGTAT 17706

Db 39 CAGCTNATGCAAAATCTCGCTCTCCAGGTTCAAGCGAT 2

RESULT 2

T24892  
ID T24892 standard; CDNA to mRNA; 100 BP.



```
SQ Sequence 108 BP; 19 A; 23 C; 28 G; 37 T;

Query Match 0.2%; Score 66.8; DB 1; Length 108;
Best Local Similarity 81.5%; Pred. No. 0.063;
Matches 88; Conservative 1; Mismatches 18; Indels 1; Gaps 1;

Oy 7604 CTGTAATCCAGCAC-TTTGGAGGCTGAGTGGATGATCACCCTGAGTTGGGAGTTTG 7662
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db 108 CTATATCCAGCACTTTTGGAGGCCAAGCAGACGGATCATCTTGAAGTCAGGAGTTGG 49

Oy 7663 AGACAGCTGCCCAACATGTTAAACCCCATGCTTACTTAAATAFACA 7710
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db 48 AGACCATCTGCCCAACAYAGGAAACCTCATCTCTACAAAAAGACA 1

RESULT 5
X12095
ID X12095 standard; DNA; 108 BP.
AC X12095;
DT 30-MAR-1999 (first entry)
DE Human biallelic polymorphic DNA fragment TIGR-A003M18a.
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medication;
KW treatment; marker; ss.
OS Homo sapiens.
PN WO9820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; U20313.
PR 06-NOV-1996; US-030455.
PA (WHEED) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
DR WPI; 98-286974/25.
PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
PS Claim 1; Page 219; 310pp; English.
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC X09121-X10268. The base occupying the polymorphic site is indicated by
CC the appropriate IUPAC-TUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC segments can also be used to produce medicaments for the treatment or
CC prophylaxis of such diseases.
SQ Sequence 108 BP; 19 A; 23 C; 28 G; 37 T;

Query Match 0.2%; Score 64.8; DB 1; Length 108;
Best Local Similarity 81.1%; Pred. No. 0.12;
Matches 86; Conservative 1; Mismatches 18; Indels 1; Gaps 1;

Oy 19552 TTTTATTTAGATAGAGTGGGTTTACCATGCTGGCCAGGCTGGTTCGAACCTCCTGACTT 19611
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db 3 TCTTTTGTAGATAGAGTGGTTCCTTGTGGCCAGGATGCTCTCGAATCCTCCTGACTT 62

Oy 19612 CAGGCGATCTGCCCGCTCAGCCCTCCCAA-CTGCTAGGATTACAG 19656
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db 63 CAAGTGATCCGCTGCTGCTTGGCCCTCCCAAAAGTGCTGGGATTATAG 108
```

```
RESULT 6
T25854
ID T25854 standard; cDNA to mRNA; 91 BP.
AC T25854;
DT 22-OCT-1996 (first entry)
DE Human gene signature HUMGS08084.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATSU) MATSUBARA K.
PA (OKUBO) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1944; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 91 BP; 18 A; 22 C; 28 G; 18 T;

Query Match 0.2%; Score 64.2; DB 1; Length 91;
Best Local Similarity 80.9%; Pred. No. 0.15;
Matches 72; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Oy 4120 ATCACTTAACTCAGGAGGCGAGGTTGTCAGTGACTGAGATCACACCATTCCTCCAG 4179
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db 2 ATCACTTGAGCCTAGGAGGCGAGNGTTTCAAGTGAGTGAGTGAGTGGCACTCCTCGCTCCAG 61

Oy 4180 CTTGGGTGACAGTGTGAGACTTCTGTCTCA 4208
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
Db 62 CTTNGGTGACAGCGTGAGANNCTGTCTCA 90

RESULT 7
T26828
ID T26828 standard; cDNA to mRNA; 108 BP.
AC T26828;
DT 14-NOV-1996 (first entry)
DE Human gene signature HUMGS09078.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATSU) MATSUBARA K.
PA (OKUBO) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
```

PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues

PS Claim 1; Page 2182; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 108 BP; 18 A; 33 C; 23 G; 28 T;

Query Match 0.2%; Score 64.4; DB 1; Length 108;  
Best Local Similarity 80.4%; Pred. No. 0.14;  
Matches 74; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 11655 GATCTCTGACCTCGTATCGCCGCCCTCAGCTCCCAAGTCTGGGATACAGGAT 11714  
||||| ||||||| ||| ||||| || ||||||| ||||||| ||||||| ||||||| ||  
Db 1 GATCTCTGACCTCGTATCGCCGCCGNTGCGCTCCCAAGTCTGGGATACAGGAT 60

QY 11715 GAGCCACTGCGCCGCCGCTCTTTTCTTTT 11746  
||||| ||||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||  
Db 61 GAGCCACCACGCCGCTGTTTATTCTTAT 92

RESULT 8

T26213 ID T26213 standard; cDNA to mRNA; 103 BP.  
AC T26213;  
DT 13-NOV-1996 (first entry)  
DE Human gene signature HUMGS08452.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN WO9514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
PI Matsubara K. Okubo K;  
DR WPI; 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues

PS Claim 1; Page 2029; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.

SQ Sequence 103 BP; 33 A; 21 C; 25 G; 23 T;

Query Match 0.2%; Score 63.6; DB 1; Length 103;  
Best Local Similarity 76.5%; Pred. No. 0.19;  
Matches 78; Conservative 0; Mismatches 24; Indels 0; Gaps 0;  
QY 10154 GATCTCTTGAGCTAGAAAGTTTGGGAGCAGTGAGTATGATTCACCTGCACCTCCA 10213  
||||| ||||||| ||| ||||||| || ||||||| ||||||| ||||||| ||||||| ||  
Db 1 GATCACTTGAGTCCAGGAGTTTGGTTACAGTGGAGTATGATGGCCACCTGCACCTCCA 60

QY 10214 GCCTGGCCAAATGCAAAATCTCTCAAAACAAAAACA 10255  
||||||| ||| ||| ||||| || ||||| || ||||| || ||||| || ||||| ||  
Db 61 GCCTGGCCACAGAGTAAGAACAATGCTCTTAAAGAAAAAAA 102

RESULT 9

T20927 ID T20927 standard; cDNA to mRNA; 103 BP.  
AC T20927;  
DT 24-JUL-1996 (first entry)  
DE Human gene signature HUMGS02180.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN WO9514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
PI Matsubara K. Okubo K;  
DR WPI; 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues

PS Claim 1; Page 758-759; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 103 BP; 22 A; 27 C; 21 G; 31 T;

Query Match 0.2%; Score 62.8; DB 1; Length 103;  
Best Local Similarity 76.0%; Pred. No. 0.24;  
Matches 76; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 17704 GATTCCTCGCTCAGCTTCCCAAGTAGCTGGGATACAGATGCACGCCACCATGCCGAG 17763  
||||| ||||| ||| ||||||| ||| ||||||| || ||||| || ||||||| ||  
Db 1 GATCCTCCCACTCCCACTCCCAAGTAGCTGGCTACAGTGTGTGCCCATGTCCAG 60

QY 17764 CTAATTTTGTATTTTGAAGAGACGGAATTCACCATG 17803  
||||| ||||||| ||| ||||| || ||||| || ||||| || ||||| || ||||| ||  
Db 61 CTGATTTTNGTATTTTNAAGTAGGACACAGATATTTCTCCATG 100

RESULT 10

T20927/c ID T20927 standard; cDNA to mRNA; 103 BP.





	Query Match	0.28;	Score 61;	DB 1;	Length 103;	
	Best Local Similarity	75.2%;	Pred. No. 0.44;			
	Matches	76;	Conservative	0;	Mismatches	25; Indels 0; Gaps 0;
QY	17606	TTCTTTTTTTTTTTCGAATAGAGTGTCGCGTCTGTACACCAGCGTGGAGNGCAGTGGCGCAA	17665			
Db	102	TTTTTTTCTTAAGACATGTTCTACTCTGTGGCCAGCGCTGGAGTGCAGTGGTGCCA	43			
QY	17666	TCTCAGTCACTGCAACGTCGCCCTCTCTGGGTTTCAAAGTGAT	17706			
Db	42	TCATAGTCACTGTACACCAAACTCTCGGACTCAAGTGAT	2			
RESULT 13						
ID	X12087/c					
AC	X12087	standard; DNA; 100 BP.				
DT	30-MAR-1999	(first entry)				
DE	Human biallelic polymorphic DNA fragment EST98276a.					
KW	polymorphism; biallelic; human; forensic; paternity testing; disease;					
KW	detection; phenotypic typing; characteristic; infection; hereditary;					
KW	autoimmune disease; cancer; inflammation; drug; therapy; medication;					
KW	treatment; marker; ss.					
OS	Homo sapiens.					
PN	WO9820165-A2.					
PD	14-MAY-1998.					
PR	05-NOV-1997; U20313.					
PR	06-NOV-1996; US-030455.					
PI	(WHED ) WHITEHEAD INST BIOMEDICAL RES.					
PT	Rudson T, Lander ES, Wang D;					
DR	WPI; 98-286974/25.					
PT	New isolated nucleic acid segments from the human genome - used for					
PT	determining polymorphic forms for use in e.g. forensics, paternity					
PT	testing or phenotypic typing for disease					
PS	Claim 1; Page 219; 31opp; English.					
CC	X12069-X12937 are human DNA fragments which contain biallelic polymorphic					
CC	markers which have been isolated using the primers represented in					
CC	X09121-X120268. The base occupying the polymorphic site is indicated by					
CC	the appropriate IUPAC-RUB ambiguity code. These fragments can be used in					
CC	methods for determining polymorphic forms in an individual for use in					
CC	e.g. forensics, paternity testing or for phenotypic typing for diseases					
CC	such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,					
CC	muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial					
CC	hypercholesterolemia, polycystic kidney disease, hereditary					
CC	spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary					
CC	haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos					
CC	syndrome, osteogenesis imperfecta, acute intermittent porphyria,					
CC	autoimmune diseases, inflammatory cancer, diseases of the nervous					
CC	system, infection by pathogenic microorganisms, and characteristics such					
CC	as longevity, appearance (e.g. baldness, obesity), strength, speed,					
CC	endurance, fertility, and susceptibility or receptivity to particular					
CC	drugs or therapeutic treatments. The isolated polymorphic nucleic acid					
CC	segments can also be used to produce medicaments for the treatment or					
CC	prophylaxis of such diseases.					
SQ	Sequence	100 BP; 21 A; 25 C; 22 G; 31 T;				
	Query Match	0.28;	Score 59.6;	DB 1;	Length 100;	
	Best Local Similarity	74.0%;	Pred. No. 0.71;			
	Matches	74;	Conservative	1;	Mismatches	25; Indels 0; Gaps 0;
QY	4913	TGTGGCTCACACCTGTTAATCCACACATTTTGGGAGGCTGAGCGGGCAGATCATCTGAGG	4972			
Db	100	TGTGACTCACACCTATTATATCTTGGCATTTTAGGAGGCTTAGAAGGAGGATGTTTGA	41			
QY	4973	TCAGAAGTTCAGACACCGCTTGGCCAAACATGGCGAAACCC	5012			
Db	40	CCAGSAGCTCAAGACCAKCTCTGGGAAACATAGCAAGACTC	1			
RESULT 14						
ID	X12085/c					
ID	X12085	standard; DNA; 100 BP.				

AC	X12085;
DE	30-MAR-1999 (first entry)
DT	Human biallelic polymorphic DNA fragment EST98276c.
DD	Polyorphism; biallelic; human; forensic; paternity testing; disease;
KW	detection; phenotypic typing; characteristic; infection; hereditary;
KW	autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW	treatment; marker; ss.
OS	Homo sapiens.
PN	WO9820165-A2.
PD	14-MAY-1998.
PF	05-NOV-1997; U20313.
PR	06-NOV-1996; US-030455.
PA	(WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI	Hudson T, Lander ES, Wang D;
DR	WPI; 98-286974/25.
PT	New isolated nucleic acid segments from the human genome - used for
PT	determining polymorphic forms for use in e.g. forensics, paternity
PT	testing or phenotypic typing for disease
PS	Claim 1; Page 218; 310pp; English.
CC	X10269-x12037 are human DNA fragments which contain biallelic polymorphic
CC	markers which have been isolated using the primers represented in
CC	X09121-x10268. The base occupying the polymorphic site is indicated by
CC	the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
CC	methods for determining polymorphic forms in an individual for use in
CC	e.g. forensics, paternity testing or for phenotypic typing for diseases
CC	such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC	mucular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC	hypercholesterolemia, polycystic kidney disease, hereditary
CC	spherocytosis, von Willibrand's disease, tuberous sclerosis, hereditary
CC	haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC	syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC	autoimmune diseases, inflammation, cancer, diseases of the nervous
CC	system, infection by pathogenic microorganisms, and characteristics such
CC	as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC	endurance, fertility, and susceptibility or receptivity to particular
CC	drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC	segments can also be used to produce medicaments for the treatment or
CC	prophylaxis of such diseases.
CC	Sequence 100 BP; 22 A; 25 C; 22 G; 30 T;
SQ	
Query Match 0.2%; Score 59.6; DB 1; Length 100;	
Best Local Similarity 74.0%; Pred. NO. 0.71;	
Matches 74; Conservative 1; Mismatches 25; Indels 0; Gaps 0;	
QY	4913 TTGTGGCTCACACCTGTAATCCAGACACTTTGGAGGCTTGAGCGGGCAGATCATCTGAGG 4972 
Db	100 TGAGTACTCACACCTAATAATCTCTGGCACTTTTGAGAGGCTTAGGAAGGAGGATTGTTTGA 41
QY	4973 TCAGAGTTCCAGACCAGCGCTGCCAACATGCGGAACCC 5012 
Db	40 CCAGGAGCTCAAGACCATCTCTGGGAACAATACAGACATCT 1
RESULT 15	
X12086/c	
ID	X12086 standard; DNA; 100 BP.
AC	X12086;
DT	30-MAR-1999 (first entry)
DD	Human biallelic polymorphic DNA fragment EST98276b.
KW	Polyorphism; biallelic; human; forensic; paternity testing; disease;
KW	detection; phenotypic typing; characteristic; infection; hereditary;
KW	autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW	treatment; marker; ss.
OS	Homo sapiens.
PN	WO9820165-A2.
PD	14-MAY-1998.
PF	05-NOV-1997; U20313.
PR	06-NOV-1996; US-030455.
PA	(WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI	Hudson T, Lander ES, Wang D;
DR	WPI; 98-286974/25.
PT	New isolated nucleic acid segments from the human genome - used for

PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease  
PS Claim 1; Page 219; 310pp; English.  
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
CC markers which have been isolated using the primers represented in  
CC X09121-X10268. The base occupying the polymorphic site is indicated by  
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases. 25 C; 22 G; 31 T;  
SQ Sequence 100 Bp; 21 A;

Query Match Score 59.6; DB 1; Length 100;  
Best Local Similarity 74.0%; Pred. No. 0.71;  
Matches 74; Conservative 1; Mismatches 25; Indels 0; Gaps 0;  
QY 4913 TGTGGCTCACACCTGTAATCCAGCAGCTTTGGAGGCTGAGCGGCGAGATCATCTGAGG 4972  
DB 100 TGTGACTCACACCTATAATCCTGGCACTTTAGGAGGCTKAGGAAGGAGGATTGTTTGA 41  
QY 4973 TCAGAGTTCAGACGAGCTGCGCAACATGCGGAACCC 5012  
DB 40 CCAGGAGCTCAAGACCATCTCTGGGAACATAGCAAGACTC 1



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2000, 23:50:22 ; Search time 8476.34 Seconds  
(without alignments)  
13867.722 Million cell updates/sec

Title: US-08-852-495C-1\_COPY\_56000\_85000  
Perfect score: 29001  
Sequence: 1 ATGAACAAGGCTGACTGAT.....CAGGAGACTAGACTTTTATT 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 156056

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : EST:  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result % Query  
SUMMARIES

No.	Score	Match	Length	DB	ID	Description
1	94.8	0.3	106	37	AA703692	ag81a10.r
2	91.4	0.3	109	30	AA243009	zr25h02.s
3	91.4	0.3	109	84	BI7434	345K2.TVB C
4	88.8	0.3	100	42	AI077628	oy26f04.s
5	88.8	0.3	101	39	AA835205	ak64h01.s
6	88.6	0.3	103	84	B48914	RPC111-4A12
7	88.2	0.3	109	84	BI7434	345K2.TVB C
8	88.2	0.3	109	94	AQ028426	CIT-HSP-2
9	88.2	0.3	110	94	AQ003188	RPC111-ID
10	87.6	0.3	110	106	AQ386882	RPC111-13
11	86.8	0.3	106	105	AQ264176	CITBI-EI-
12	86.6	0.3	110	30	AA244245	nc07a04.s
13	85.8	0.3	107	24	H67040	yu68c01.r1
14	85.8	0.3	108	84	B65160	CIT-HSP-201
15	85.2	0.3	106	38	AA812141	OB48h02.s
16	85.2	0.3	110	39	AA897366	am06h02.s
17	84.6	0.3	107	35	AA565533	nk42b11.s
18	84.6	0.3	103	108	AQ352444	RPCI-11-3
19	84.4	0.3	103	38	AA807640	nx08b05.s
20	84.2	0.3	105	105	AQ282107	RPC111-94
21	84.4	0.3	110	106	AQ386882	RPC111-13
22	83.8	0.3	103	94	AQ028649	CIT-HSP-2
23	83.8	0.3	103	108	AQ352444	RPCI-11-3
24	83.6	0.3	106	63	AI991750	wt48e01.x
25	83.6	0.3	107	33	AA385808	EST99495
26	84	0.3	109	30	AA243009	zr25h02.s
27	83.4	0.3	101	35	AA583697	nn58f10.s
28	83.6	0.3	106	63	AI991750	wt48e01.x
29	83	0.3	91	38	AA780764	ac68f12.s
30	83.2	0.3	105	30	AA218889	zg15d04.s
31	83.2	0.3	107	33	AA385808	EST99495
32	82.8	0.3	103	108	AQ382186	RPCI-11-4
33	82.8	0.3	103	108	AQ382186	RPCI-11-4
34	83	0.3	107	35	AA565533	nk42b11.s
35	82.4	0.3	101	35	AA583697	nn58f10.s
36	82.6	0.3	105	28	AA078003	7H12D08 C
37	82.8	0.3	105	109	AQ637292	RPCI-11-4
38	82.8	0.3	110	109	AQ634950	RPCI-11-4
39	82.2	0.3	103	108	AQ334922	RPCI-11-3
40	82	0.3	106	105	AQ264176	CITBI-EI-
41	82.2	0.3	106	105	AQ282340	RPC111-80
42	82	0.3	106	108	AQ544957	CITBI-EI-
43	81.8	0.3	109	94	AQ028426	CIT-HSP-2
44	81.8	0.3	109	105	AQ265749	CITBI-EI-
45	81.4	0.3	99	34	AA486800	abl9a06.r

## ALIGNMENTS

RESULT 1  
AA703692  
LOCUS ag81a10.r1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone  
DEFINITION IMAGE:1140858 5' similar to contains Alu repetitive element; mRNA  
sequence.  
ACCESSION AA703692 106 bp mRNA  
VERSION AA703692.1 GI:2713610  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 106)  
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.  
TITLE WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397630.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28ml3 rev1 ET from Amersham  
High quality sequence stop: 53.  
Location/Qualifiers  
1. 106  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1140858"  
/clone\_lib="Stratagene hNT neuron (#937233)"  
/dev\_stage="hNT neurons"  
/lab\_host="SOLR (kanamycin resistant)"  
/notes="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; Cloned unidirectionally. Primer: Oligo dt.  
Differentiated, post mitotic hNT neurons. Average insert  
size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'  
GAATCGGCACGAG 3' -3' adaptor sequence: 5'  
CTCGAGTTTTTTTTTTTTTTT 3' "

FEATURES  
source

BASE COUNT 19 a 29 c 29 g 29 t  
ORIGIN  
Query Match 0.3%; Score 94.8; DB 37; Length 106;  
Best Local Similarity 93.04%; Pred. No. 0.044;  
Matches 99; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 11610 TTTTACTAGATACGGGTTTCACTTGTGTTAACACGAGTGTCTCGATCTCTGACCTCG 11669  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1 TTTTGTAGTAGACGAGGTTTCCCGTGTAGCCGAGGATGCTCGATCTCTGACCTCG 60  
QY 11670 TGATCGGCCCTCCAGCTCCCAAGTCTGGGATTCACGAGTG 11715  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 61 TGATCGGCCCTCCAGCTCCCAAGTCTGGGATTCACGAGTG 106  
RESULT 2  
AA243009 109 bp mRNA EST 11-MAR-1998  
LOCUS zr25h02.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens  
DEFINITION cDNA clone IMAGE:66467 3' similar to contains Alu repetitive  
element; contains element LTR1 repetitive element ;, mRNA sequence.  
ACCESSION AA243009  
VERSION AA243009.1 GI:1873869  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 109)  
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.  
TITLE WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT On Dec 3, 1996 this sequence version replaced gi:1126869.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert length: 1127 Std Error: 0.00  
Seq primer: -41ml3 fwd. ET from Amersham  
High quality sequence stop: 102.

FEATURES  
source

Location/Qualifiers  
1. .109  
/organism="Homo sapiens"  
/db\_xref="GDB:5426481"  
/db\_xref="taxon:9606"  
/clone="IMAGE:664467"  
/clone\_lib="Stratagene NT2 neuronal precursor 937230"  
/type="neuroepithelial cells"  
/dev\_stage="Ntera-2 neuroepithelial cells"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="organ: brain; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Uninduced, exponentially growing neuroepithelial  
cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGCAGG  
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3"

BASE COUNT 19 a 30 c 30 g 30 t  
ORIGIN

Query Match 0.3%; Score 91.4; DB 30; Length 109;  
Best Local Similarity 89.9%; Pred. No. 0.11; Mismatches 0; Gaps 0;  
Matches 98; Conservative 0; Indels 0; Gaps 0;

QY 11607 GATTTTATAGATACGGGGTTTCACCTTGTAAACAGGATGCTCGATCTCCTGACC 11666  
||||| 11607 GATTTTATAGATACGGGGTTTCACCTTGTAAACAGGATGCTCGATCTCCTGACC 11666  
Db 1 GATTTTATAGATACGGGGTTTCACCTTGTAAACAGGATGCTCGATCTCCTGACC 11666

QY 11667 TCGTATCGCGCGCTCAGCCTCCCAAGTCTGGGATTACAGGAGTG 11715  
||||| 11667 TCGTATCGCGCGCTCAGCCTCCCAAGTCTGGGATTACAGGAGTG 11715  
Db 61 TCGTATCGCGCGCTCAGCCTCCCAAGTCTGGGATTACAGGAGTG 11715

RESULT 3  
B17434/c

LOCUS B17434 109 bp DNA GSS 04-JUN-1998  
DEFINITION 345K2.TVB C1978SKA1 Homo sapiens genomic clone A-345K02, genomic survey sequence.  
ACCESSION B17434  
VERSION B17434.1 GI:2125183  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 109)  
AUTHORS Adams, M.D., Kelley, J.M., Rounsley, S.R. and Venter, J.C.  
TITLE Use of a BAC End Sequence Database for Sequence-Ready Map Building  
JOURNAL Unpublished (1997)  
COMMENT Other GSSs: 345K02.TP 345K02.TPB  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html  
Seq primer: 77  
Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
1. .109  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="A-345K02"  
/clone\_lib="C1978SKA1"  
/sex="Female"  
/cell\_type="Fibroblast"  
/note="Vector: pBAC108L; Site\_1: HindIII; Site\_2: HindIII;  
Caltech Human BAC Library A1"

BASE COUNT 24 a 30 c 31 g 24 t

ORIGIN

Query Match 0.3%; Score 91.4; DB 84; Length 109;  
Best Local Similarity 89.9%; Pred. No. 0.11; Mismatches 0; Gaps 0;  
Matches 98; Conservative 0; Indels 0; Gaps 0;

QY 19558 TAGTAGAGATGGGTTTACCATCTGGCCAGGCTGCTCGAACCTCTGACCTCAGGCG 19617  
||||| 19558 TAGTAGAGATGGGTTTACCATCTGGCCAGGCTGCTCGAACCTCTGACCTCAGGCG 19617  
Db 109 TAGTTGACAGGGGTTTACCATCTGGCCAGGCTGCTCGAACCTCTGACCTCAGGCG 50  
||||| 109 TAGTTGACAGGGGTTTACCATCTGGCCAGGCTGCTCGAACCTCTGACCTCAGGCG 50

QY 19618 ATCTGCGCCCTCAGCTCCCAAGTCTAGGATTACAGCGCTGAGGCCA 19666  
||||| 19618 ATCTGCGCCCTCAGCTCCCAAGTCTAGGATTACAGCGCTGAGGCCA 19666  
Db 49 ATCCGCCCATCATCAGCTCCCAAGTCTAGGATTATAGTATGATGAGGCCA 1  
||||| 49 ATCCGCCCATCATCAGCTCCCAAGTCTAGGATTATAGTATGATGAGGCCA 1

RESULT 4  
AI077628/c  
LOCUS AI077628 100 bp mRNA EST 24-SEP-1998  
DEFINITION oy26f04.s1 Soares senescent fibroblasts\_NbHSF Homo sapiens CDNA (HUMAN);, mRNA sequence.  
ACCESSION AI077628.1 GI:3412036  
VERSION AI077628  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 100)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2153443.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Trace considered overall poor quality  
Insert Length: 820 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers  
1. .100  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1666975"  
/clone\_lib="Soares senescent fibroblasts\_NbHSF"  
/tissue\_type="senescent fibroblast"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pT7T3D (Pharmacia) with a modified polylinker V\_TYPE: phagemid; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5'  
TCTTACCAATCTCAAGTCGGACGCCGCGCATTTTTTTTTTTTTTTT 3']  
double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 15 a 29 c 28 g 28 t

ORIGIN

Query Match 0.3%; Score 88.8; DB 42; Length 100;  
Best Local Similarity 93.0%; Pred. No. 0.22; Mismatches 0; Gaps 0;  
Matches 93; Conservative 0; Indels 0; Gaps 0;

QY 22054 AGAGCCTGGTACGAGGCGCACTCTGGTCAACGAAAGGACCGCTGCTTGGCTCC 22113  
||||| 22054 AGAGCCTGGTACGAGGCGCACTCTGGTCAACGAAAGGACCGCTGCTTGGCTCC 22113





AUTHORS Adams,M.D., Kelley,J.M., Rounsley,S.R. and Venter,J.C.  
 TITLE Use of a BAC End Sequence Database for Sequence-Ready Map Building  
 JOURNAL Unpublished (1997)  
 COMMENT Other.GSSs: 345K02.TP 345K02.TPB  
 Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadams@tigr.org  
 Clones are available from Research Genetics (info@resgen.com). BAC  
 end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
 Seq primer: T7  
 Class: BAC ends.

FEATURES Location/Qualifiers  
 source 1..109  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="A-345K02"  
 /clone\_lib="CIT978SKA1"  
 /sex="Female"  
 /cell\_type="Fibroblast"  
 /note="Vector: pBAC108L; Site\_1: HindIII; Site\_2: HindIII;  
 CalTech Human BAC Library A1"  
 BASE COUNT 24 a 30 c 31 g 24 t  
 ORIGIN

Query Match 0.3%; Score 88.2; DB 84; Length 109;  
 Best Local Similarity 88.1%; Pred. No. 0.25;  
 Matches 96; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 QY 7594 TGGCTCAGCTGTAAATCCAGCACTTTGGGAGGCTGGAGTGATGCATCAGGT 7653  
 Db 1 TGGCTCATACCTAATATCTAGCACTTTGGGAGGCTGATGTGGCGGATCACTGAGGTC 60  
 QY 7654 GGGCTTTGAGACCACTGGCCACATGGTAAACCCCATCTACTA 7702  
 Db 61 GGGAGTTGAGACCACTGGCCACCNCTGGTGAACCCCGTCTCAACTA 109

RESULT 8  
 LOCUS A0028426 109 bp DNA GSS 30-JUN-1998  
 DEFINITION CIT-HSP-2313G15.TF CIT-HSP Homo sapiens genomic clone 2313G15,  
 genomic survey sequence.  
 ACCESSION A0028426  
 VERSION A0028426.1 GI:3268648  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 109)  
 AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,  
 Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,  
 Simon,M. and Venter,J.C.  
 TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map  
 Building (1998)  
 JOURNAL Unpublished (1998)  
 COMMENT Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadams@tigr.org  
 Clones are available from Research Genetics (info@resgen.com). BAC  
 end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
 Seq primer: M13-21

Class: BAC ends.  
 Location/Qualifiers  
 source 1..109  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="2313G15"  
 /clone\_lib="CIT-HSP"  
 /sex="Male"  
 /cell\_type="Sperm"  
 /note="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2:  
 HindIII"  
 BASE COUNT 19 a 36 c 25 g 29 t  
 ORIGIN

Query Match 0.3%; Score 88.2; DB 94; Length 109;  
 Best Local Similarity 88.1%; Pred. No. 0.25;  
 Matches 96; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 QY 17610 TTTTGTGAAATAGAGTCTCGCTCTGTCAACCCAGGCTGGAGTGGCGCAATCTC 17659  
 Db 1 TGTGTTCTGAGAGGACTCTCAGTCTGTCAACCCAGGCTGGAGTGGCGACAGTCTG 60  
 QY 17670 AGCTCACTGCAACGTCGCCCTCTGGGTTCAGTGAATCTCTCGCTCA 17718  
 Db 61 AGCTCACTGCAACCTCCACCTCTCTGGTTCAAGCGATTCTCTCGCTCA 109

RESULT 9  
 LOCUS A0003188 110 bp DNA GSS 14-APR-1999  
 DEFINITION RPC111-1D10.TPN RPCI-11 Homo sapiens genomic clone RPCI-11-1D10,  
 genomic survey sequence.  
 ACCESSION A0003188  
 VERSION A0003188.1 GI:3030392  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 110)  
 AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,  
 Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and  
 Venter,J.C.  
 TITLE Use of BAC End Sequences for Sequence-Ready Map Building (1998)  
 JOURNAL Unpublished (1998)  
 COMMENT Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadams@tigr.org  
 Clones are derived from the human BAC library RPCI-11. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from  
 Research Genetics (info@resgen.com). BAC end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES Location/Qualifiers  
 source 1..110  
 /organism="Homo sapiens"  
 /db\_xref="GDB:7500081"  
 /db\_xref="taxon:9606"  
 /clone="RPCI-11-1D10"  
 /clone\_lib="RPCI-11"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
 RPC111 Human Male BAC Library"  
 BASE COUNT 22 a 27 c 26 g 35 t

```
ORIGIN
Query Match      0.3%; Score 88.2; DB 94; Length 110;
Best Local Similarity 88.1%; Pred. No. 0.25; 13; Indels 0; Gaps 0;
Matches 96; Conservative 0; Mismatches 0;

Qy 11602 TTTTGTATTTTGTAGTACGCGGTTTCACTTGTGTTAAACGAGGATGCTCGATCTCC 11661
||||| 11602 TTTTGTATTTTGTAGTACGCGGTTTCACTTGTGTTAAACGAGGATGCTCGATCTCC 11661
Db 2 TTTTGTATTTTGTAGTACGCGGTTTCACTTGTGTTAAACGAGGATGCTCGATCTCT 61
||||| 2 TTTTGTATTTTGTAGTACGCGGTTTCACTTGTGTTAAACGAGGATGCTCGATCTCT 61

Qy 11662 TGACCTCGTGATCGCGCCGCTCAACCTCCCAAAAGTGTGGGATTACAG 11710
||||| 11662 TGACCTCGTGATCGCGCCGCTCAACCTCCCAAAAGTGTGGGATTACAG 11710
Db 62 TGACCTCATGATCCACCTCGCGCAGCCCTCCCAAAAGTGTGGGATTACAG 110
||||| 62 TGACCTCATGATCCACCTCGCGCAGCCCTCCCAAAAGTGTGGGATTACAG 110

RESULT 10
AQ386882/c 110 bp DNA GSS 21-MAY-1999
LOCUS
DEFINITION RPC111-13414.TV RPCI-11 Homo sapiens genomic clone RPCI-11-13414,
genomic survey sequence.
ACCESSION AQ386882
VERSION AQ386882.1 GI:4357905
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 110)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
Venter,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPC111-13414.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pletredejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.
FEATURES
Source Location/Qualifiers
1..110
/organism="Homo sapiens"
/db_xref="GDB:7551267"
/db_xref="taxon:9606"
/clone_lib="RPCI-11"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/notes="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"
BASE COUNT 26 a 26 c 38 g 20 t
ORIGIN

Query Match      0.3%; Score 87.6; DB 106; Length 110;
Best Local Similarity 87.3%; Pred. No. 0.29;
Matches 96; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 19569 GGGTTTACCATGTCGGCAGCGTGTCTCGAATCTCTGACCTCAGCGATCTGCCGCC 19628
||||| 19569 GGGTTTACCATGTCGGCAGCGTGTCTCGAATCTCTGACCTCAGCGATCTGCCGCC 19628
Db 110 GGGTTTACCATGTCGGCAGCGTGTCTGACTCTCTGACCTCAGCGATCTGCCGCC 51
||||| 110 GGGTTTACCATGTCGGCAGCGTGTCTGACTCTCTGACCTCAGCGATCTGCCGCC 51

ORIGIN
Query Match      0.3%; Score 86.8; DB 105; Length 106;
Best Local Similarity 88.7%; Pred. No. 0.37; 12; Indels 0; Gaps 0;
Matches 94; Conservative 0; Mismatches 12;

Qy 11625 GGGTTTCACTTTTGTAAACGAGGATGCTCGATCTCTGACCTGATCGCGCGCTC 11684
||||| 11625 GGGTTTCACTTTTGTAAACGAGGATGCTCGATCTCTGACCTGATCGCGCGCTC 11684
Db 106 GGGTTTCACTTTTGTAAACGAGGATGCTCGATCTCTGACCTGATCGCGCGCTC 47
||||| 106 GGGTTTCACTTTTGTAAACGAGGATGCTCGATCTCTGACCTGATCGCGCGCTC 47

Qy 11685 AGCTTCCCAAGTGTGGGATTACAGGATGAGGACCTGCGCCCGG 11730
||||| 11685 AGCTTCCCAAGTGTGGGATTACAGGATGAGGACCTGCGCCCGG 11730
Db 46 GGTCTCCCAAGTGTGGGATTACAGGCGTGAGACTCTGCGCCCGG 1
||||| 46 GGTCTCCCAAGTGTGGGATTACAGGCGTGAGACTCTGCGCCCGG 1

RESULT 12
AA244245 110 bp mRNA EST 20-AUG-1997
LOCUS
DEFINITION nc07a04.s1 NCI_CGAP_Prl Homo sapiens cDNA clone IMAGE:1007406
similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION AA244245
VERSION AA244245.1 GI:1875104
KEYWORDS EST.
SOURCE human.
```

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 110)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 24, 1995 this sequence version replaced gi:534306.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1350  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuauqui,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: David B. Krizman, Ph.D.  
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/dbp/Image/Image.html](http://www-bio.llnl.gov/dbp/Image/Image.html)

Seq primer: -41m13 fwd. ET from Amersham  
High quality sequence stop: 90.

## FEATURES

source  
Location/Qualifiers

1..110  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1007406"  
/clone\_lib="NCI-CGAP\_Prl"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"  
/note="Vector: PAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from  
5,000-10,000 microdissected, histologically normal  
prostate epithelial cells. Double-stranded cDNA was  
ligated to EcoRI adaptors, 5 cycles of PCR applied to the  
cDNA with an adaptor-specific primer, and the resulting  
PCR product subcloned into PAMP10 by the UDG-cloning  
method (Life Technologies). Average insert size is 600  
bp. NOTE: Not directionally cloned. This library was  
constructed by David Krizman."

BASE COUNT 17 a 26 c 28 g 38 t 1 others

Query Match 0.3%; Score 86.6; DB 30; Length 110;  
Best Local Similarity 86.4%; Pred. No. 0.38;  
Matches 95; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 17609 TTTTCTTTTGAATAGATCTCGCTCTACCCAGGCTGAGTGCAGTGGCGAATCT 17668

Db 1 TTTTCTTTTGAATAGATCTCTGATCTTGCCAGGCTGGAGTGCAGTGGCGAANTCT 60

QY 17669 CAGCTCAGTCGACGCGCCCTCTCGGTTCAAGTCATCTCCTCGCTCA 17718

Db 61 TGGCTCAGTCGACACCTCTGCCCTCTCGGTTCAAGAGATTCTCTGCCTCA 110

## RESULT 13

LOCUS H67040 107 bp mRNA EST 27-OCT-1995  
DEFINITION yu68c01.r1 Weizmann Olfactory Epithelium Homo sapiens cDNA clone  
IMAGE:238944 5' similar to contains Alu repetitive element.; mRNA  
sequence.  
ACCESSION H67040  
VERSION H67040.1 GI:1025780  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 107)  
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chissoe, S., Dietrich, N., Dubuque, T., Pavello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
and Marra, M.

TITLE Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL Genome Res. 6 (9), 807-828 (1996)  
MEDLINE 97044478

COMMENT On Nov 29, 1993 this sequence version replaced gi:429999.

Contact: Willson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

High quality sequence stops: 101

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

Seq primer: M13RPI

High quality sequence stop: 101.

## FEATURES

Location/Qualifiers  
1..107  
/organism="Homo sapiens"  
/db\_xref="GDB:3864328"  
/db\_xref="taxon:9606"  
/clone="IMAGE:238944"  
/clone\_lib="Weizmann Olfactory Epithelium"  
/sex="Female"

/tissue\_type="olfactory epithelium"

/dev\_stage="35 year old"

/lab\_host="SOLR cells (kanamycin resistant)"

/note="Organ: nose; Vector: pBluescript SK-; Site\_1:

EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:

Oligo dT. Olfactory epithelium, normal. Average insert

size: 0.8 kb; Uni-ZAP XR Vector. Library constructed by N.

Walker, D. Lancet, Weizmann Institute of Science. -5'

adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor

sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'

24 a 37 c 20 g 24 t 2 others

## ORIGIN

Query Match 0.3%; Score 85.8; DB 24; Length 107;  
Best Local Similarity 86.9%; Pred. No. 0.48;  
Matches 93; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 10997 TAGCTGGCGGTGGTGACATGCTCTAGTCCAGCTACTGGGAGGCTGAGCAGGAGA 11056

Db 107 TAGCTGGGTGTGTAGACATGCTCTGATTCNAGCTACTCAGNAGGCTGAGTAGGAGA 48

QY 11057 ATTGCTTGAACCTCGGAGCGGAGTTGCAGTCAGCGCAGATTGCGC 11103

Db 47 ATCGCTTGAACCCAGGAGGCTTGCAGTCAGCTGAGATTGTGC 1

## RESULT 14

LOCUS B65160 108 bp DNA GSS 21-JUN-1998  
DEFINITION CIT-HSP-2017G2.TRB CIT-HSP Homo sapiens genomic clone 2017G2,  
genomic survey sequence.  
ACCESSION B65160  
VERSION B65160.1 GI:2639138  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 108)

AUTHORS Adams, M.D., Rounsley, S.D., Field, C.E., Bass, S., Linher, K.,



---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2000, 04:56:24 ; Search time 372.51 Seconds  
(without alignments)  
10119.719 Million cell updates/sec

Title: US-08-852-495c-1\_copy\_56000\_85000  
Perfect score: 29001  
Sequence: 1 ATGACCAAGGCTGACTGAT.....CAGGAGACTAGAGTTTATT 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 374504

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83.2	0.3	105	4	US-08-481-658B-65
2	83.2	0.3	105	4	US-08-477-504A-65
3	83.2	0.3	105	4	US-08-486-756A-65
4	83.2	0.3	105	4	US-08-485-862B-65
5	83.2	0.3	105	5	US-08-787-739-65
6	81.6	0.3	105	4	US-08-481-658B-65
7	81.6	0.3	105	4	US-08-477-504A-65
8	81.6	0.3	105	4	US-08-486-756A-65
9	81.6	0.3	105	4	US-08-485-862B-65
10	81.6	0.3	105	5	US-08-787-739-65
11	67.2	0.2	84	3	US-08-454-557C-91
12	67.2	0.2	84	4	US-08-340-426D-91
13	67.2	0.2	84	4	US-08-450-673C-91
14	67.2	0.2	84	6	PCT-US95-1711A-91
15	62.2	0.2	84	3	US-08-454-557C-91
16	62.2	0.2	84	4	US-08-340-426D-91
17	62.2	0.2	84	4	US-08-450-673C-91
18	62.2	0.2	84	6	PCT-US95-1711A-91
19	58.8	0.2	78	3	US-08-454-557C-70
20	58.8	0.2	78	4	US-08-340-426D-70
21	58.8	0.2	78	4	US-08-450-673C-70
22	58.8	0.2	78	6	PCT-US95-1711A-70
23	54.6	0.2	85	3	US-08-454-557C-92
24	54.4	0.2	85	3	US-08-454-557C-92
25	54.6	0.2	85	4	US-08-340-426D-92
26	54.4	0.2	85	4	US-08-340-426D-92
27	54.6	0.2	85	4	US-08-450-673C-92

c	28	54.4	0.2	85	4	US-08-450-673C-92	Sequence 92, Appl
c	29	54.6	0.2	85	6	PCT-US95-1711A-92	Sequence 92, Appl
c	30	54.4	0.2	85	6	PCT-US95-1711A-92	Sequence 92, Appl
c	31	53.6	0.2	60	3	US-08-454-557C-60	Sequence 60, Appl
c	32	53.6	0.2	60	4	US-08-340-426D-60	Sequence 60, Appl
c	33	53.6	0.2	60	4	US-08-450-673C-60	Sequence 60, Appl
c	34	53.6	0.2	60	6	PCT-US95-1711A-60	Sequence 60, Appl
c	35	53.2	0.2	76	3	US-08-454-557C-69	Sequence 69, Appl
c	36	53.2	0.2	76	4	US-08-340-426D-69	Sequence 69, Appl
c	37	53.2	0.2	76	4	US-08-450-673C-69	Sequence 69, Appl
c	38	53.2	0.2	76	6	PCT-US95-1711A-69	Sequence 69, Appl
c	39	53.2	0.2	83	4	US-08-481-658B-66	Sequence 66, Appl
c	40	53.2	0.2	83	4	US-08-477-504A-66	Sequence 66, Appl
c	41	53.2	0.2	83	4	US-08-486-756A-66	Sequence 66, Appl
c	42	53.2	0.2	83	4	US-08-485-862B-66	Sequence 66, Appl
c	43	53.2	0.2	83	5	US-08-787-739-66	Sequence 66, Appl
c	44	52.6	0.2	78	3	US-08-454-557C-70	Sequence 70, Appl
c	45	52.6	0.2	78	4	US-08-340-426D-70	Sequence 70, Appl

ALIGNMENTS

RESULT 1  
US-08-481-658B-65  
; Sequence 65, Application US/08481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,658B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REFERENCE/DOCKET NUMBER: D-0021.3E  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-481-658B-65

Query Match 0.3%; Score 83.2; DB 4; Length 105;  
Best Local Similarity 87.5%; Pred. No. 1.3e-09;

Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 27998 TTTTGTATTTTATTAGACAGGGTTTCACTATGTTGGCCAGGCTGATCTCAAACCTCC 28057

Db 2 TTTTACATCTTTAGACAGGGTTTCACTATGTTGGCCAGGCTGATCTCAAACCTCC 61

Qy 28058 TGACCTCATGATCCGCCCTGCGCTTCTCAAAGTGTCTGGGAT 28101

Db 62 TGACCTGTGTATCCACAGCCTCGGCCCTCCCAAAAGTGTCTGGGAT 105

## RESULT 2

US-08-477-504A-65

; Sequence 65, Application US/08477504A

; Patent No. 5972353

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

; CITY: Tiburon

; STATE: California

; COUNTRY: USA

; ZIP: 94920

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/477,504A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/260,190

; FILING DATE: 15-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lauder, Leona L.

; REGISTRATION NUMBER: 30,863

; REFERENCE/DOCKET NUMBER: D-0021.3D

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-435-2034

; TELEFAX: 415-435-0727

; INFORMATION FOR SEQ ID NO: 65:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 105 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-477-504A-65

Query Match 0.3%; Score 83.2; DB 4; Length 105;

Best Local Similarity 87.5%; Pred. No. 1.3e-09;

Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 27998 TTTTGTATTTTATTAGACAGGGTTTCACTATGTTGGCCAGGCTGATCTCAAACCTCC 28057

Db 2 TTTTACATCTTTAGACAGGGTTTCACTATGTTGGCCAGGCTGATCTCAAACCTCC 61

Qy 28058 TGACCTCATGATCCGCCCTGCGCTTCTCAAAGTGTCTGGGAT 28101

Db 62 TGACCTGTGTATCCACAGCCTCGGCCCTCCCAAAAGTGTCTGGGAT 105

## RESULT 3

US-08-486-756A-65

; Sequence 65, Application US/08486756A

; Patent No. 5981711

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

; CITY: Tiburon

; STATE: California

; COUNTRY: USA

; ZIP: 94920

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/486,756A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/260,190

; FILING DATE: 15-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lauder, Leona L.

; REGISTRATION NUMBER: 30,863

; REFERENCE/DOCKET NUMBER: D-0021.3C

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-435-2034

; TELEFAX: 415-435-0727

; INFORMATION FOR SEQ ID NO: 65:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 105 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-486-756A-65

Query Match 0.3%; Score 83.2; DB 4; Length 105;

Best Local Similarity 87.5%; Pred. No. 1.3e-09;

Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 27998 TTTTGTATTTTATTAGACAGGGTTTCACTATGTTGGCCAGGCTGATCTCAAACCTCC 28057

Db 2 TTTTACATCTTTAGACAGGGTTTCACTATGTTGGCCAGGCTGATCTCAAACCTCC 61

Qy 28058 TGACCTCATGATCCGCCCTGCGCTTCTCAAAGTGTCTGGGAT 28101

Db 62 TGACCTGTGTATCCACAGCCTCGGCCCTCCCAAAAGTGTCTGGGAT 105

## RESULT 4

US-08-485-862B-65

; Sequence 65, Application US/08485862B

; Patent No. 5989838

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court



CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,862B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-485-862B-65

Query Match 0.3%; Score 83.2; DB 4; Length 105;  
Best Local Similarity 87.5%; Pred. No. 1.3e-09;  
Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 27998 TTTTGTATTTTATTAGACAGGTTTCACACTGTGGCCAGGCTGATCTCAAACTCC 28057  
Db 2 TTTTACATCTTTAGTAGACAGGTTTCACCATATTGGCCAGGCTGCTCTCAAACTCC 61  
QY 28058 TCACCTCATGATCCGCCCTGCCCTCAAGTCTCAAAAGTCTGGGAT 28101  
Db 62 TGACCTTGATCCACGCTCGGCCCTCCCAAAAGTCTGGGAT 105

Query Match 0.3%; Score 83.2; DB 5; Length 105;  
Best Local Similarity 87.5%; Pred. No. 1.3e-09;  
Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 27998 TTTTGTATTTTATTAGACAGGTTTCACACTGTGGCCAGGCTGATCTCAAACTCC 28057  
Db 2 TTTTACATCTTTAGTAGACAGGTTTCACCATATTGGCCAGGCTGCTCTCAAACTCC 61  
QY 28058 TCACCTCATGATCCGCCCTGCCCTCAAGTCTCAAAAGTCTGGGAT 28101  
Db 62 TGACCTTGATCCACGCTCGGCCCTCCCAAAAGTCTGGGAT 105

RESULT 5  
US-08-787-739-65  
; Sequence 65, Application US/08787739  
; Patent No. 6027887  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 96  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 369 Pine Street, Suite 610  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/787,739  
FILING DATE: 24-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,049  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/486,756  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/481,658  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,862  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,863  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/487,077  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-981-2034  
TELEFAX: 415-981-0332  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-787-739-65

Query Match 0.3%; Score 83.2; DB 5; Length 105;  
Best Local Similarity 87.5%; Pred. No. 1.3e-09;  
Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 27998 TTTTGTATTTTATTAGACAGGTTTCACACTGTGGCCAGGCTGATCTCAAACTCC 28057  
Db 2 TTTTACATCTTTAGTAGACAGGTTTCACCATATTGGCCAGGCTGCTCTCAAACTCC 61  
QY 28058 TCACCTCATGATCCGCCCTGCCCTCAAGTCTCAAAAGTCTGGGAT 28101  
Db 62 TGACCTTGATCCACGCTCGGCCCTCCCAAAAGTCTGGGAT 105

RESULT 6  
US-08-481-658B-65/c  
; Sequence 65, Application US/08481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA: US/08/481.658B  
APPLICATION NUMBER: US/08/481.658B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-481-658B-65

Query Match 0.3%; Score 81.6; DB 4; Length 105;  
Best Local Similarity 86.5%; Pred. No. 3e-09;  
Matches 90; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 25018 ATCCGAGCTCTTTGGAGGCGCTAGCGGTGATCAGGAGTTCAGACGAGC 25077  
Db 105 ATCCGAGCACTTTGGAGGCGCGAGCTGGTGGATCACAAGGTCAGGAGTTTGAGAGCAGC 46  
Qy 25078 CTCGCCAAGATGGTGAATCCGCTCTACTATAAAAGTATAAAA 25121  
Db 45 CTGGCCATATGGTGAACCCCTGCTCTACTATAAAGATGTAAAAA 2

RESULT 7  
US-08-477-504A-65/C  
Sequence 65, Application US/08/477504A  
Patent No. 5972353  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477.504A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:

NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-477-504A-65

Query Match 0.3%; Score 81.6; DB 4; Length 105;  
Best Local Similarity 86.5%; Pred. No. 3e-09;  
Matches 90; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 25018 ATCCGAGCTCTTTGGAGGCGCTAGCGGTGATCAGGAGTTCAGACGAGC 25077  
Db 105 ATCCGAGCACTTTGGAGGCGCGAGCTGGTGGATCACAAGGTCAGGAGTTTGAGAGCAGC 46  
Qy 25078 CTCGCCAAGATGGTGAATCCGCTCTACTATAAAAGTATAAAA 25121  
Db 45 CTGGCCATATGGTGAACCCCTGCTCTACTATAAAGATGTAAAAA 2

RESULT 8  
US-08-486-756A-65/C  
Sequence 65, Application US/08486756A  
Patent No. 5981711  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486.756A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

Query Match	0.3%	Score 81.6;	DB 4;	Length 105;
Best Local Similarity	86.5%	Pred. No. 3e-09;		
Matches 90;	Conservative	0;	Mismatches 14;	Indels 0;
<p>Query Match 0.3% Score 81.6; DB 4; Length 105;</p> <p>Best Local Similarity 86.5%; Pred. No. 3e-09;</p> <p>Matches 90; Conservative 0; Mismatches 14; Indels 0; Gaps 0;</p>				
QY 25018	ATCCCAAGTCTTTGGGAGGCTAGCGGCTGGATCAGAGGTCTCAGGAGTTCAGAGCAGC	25077		
Db 105	ATCCCAAGTCTTTGGGAGGCTAGCGGCTGGATCAGAGGTCTCAGGAGTTCAGAGCAGC	46		
QY 25078	CTCCCAAGTCTTTGGGAGGCTAGCGGCTGGATCAGAGGTCTCAGGAGTTCAGGAGTTCAGGAGCAGC	25121		
Db 45	CTGCGCAATATGGTGAATCCCGTCTCTACTATAAGATATAAAAA	2		
<p>RESULT 9</p> <p>US-08-485-862B-65/c</p> <p>Sequence 65, Application US/08485862B</p> <p>Patent No. 5989838</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Zavada, Jan</p> <p>APPLICANT: Pastorekova, Silvia</p> <p>APPLICANT: Pastorek, Jaromir</p> <p>TITLE OF INVENTION: MN Gene and Protein</p> <p>NUMBER OF SEQUENCES: 86</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: Leona L. Lauder</p> <p>STREET: 6 Mariposa Court</p> <p>CITY: Tiburon</p> <p>STATE: California</p> <p>COUNTRY: USA</p> <p>ZIP: 94920</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Floppy disk</p> <p>COMPUTER: IBM PC compatible</p> <p>OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/08/485,862B</p> <p>FILING DATE: 07-JUN-1995</p> <p>CLASSIFICATION: 435</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/477,504</p> <p>FILING DATE: 07-JUN-1995</p> <p>APPLICATION NUMBER: US 08/260,190</p> <p>FILING DATE: 15-JUN-1994</p> <p>ATTORNEY/AGENT INFORMATION:</p> <p>NAME: Lauder, Leona L.</p> <p>REGISTRATION NUMBER: 30,863</p> <p>REFERENCE/DOCKET NUMBER: D-0021.3D</p> <p>TELECOMMUNICATION INFORMATION:</p> <p>TELEPHONE: 415-435-2034</p> <p>TELEFAX: 415-435-0727</p> <p>INFORMATION FOR SEQ ID NO: 65:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 105 base pairs</p> <p>TYPE: nucleic acid</p> <p>STRANDEDNESS: single</p> <p>TOPOLOGY: linear</p> <p>MOLECULE TYPE: DNA (genomic)</p> <p>HYPOTHETICAL: NO</p> <p>ANTI-SENSE: NO</p> <p>US-08-485-862B-65</p>				
QY 25018	ATCCCAAGTCTTTGGGAGGCTAGCGGCTGGATCAGAGGTCTCAGGAGTTCAGAGCAGC	25077		
Db 105	ATCCCAAGTCTTTGGGAGGCTAGCGGCTGGATCAGAGGTCTCAGGAGTTCAGAGCAGC	46		
QY 25078	CTCCCAAGTCTTTGGGAGGCTAGCGGCTGGATCAGAGGTCTCAGGAGTTCAGGAGTTCAGGAGCAGC	25121		
Db 45	CTGCGCAATATGGTGAATCCCGTCTCTACTATAAGATATAAAAA	2		
<p>Query Match 0.3% Score 81.6; DB 4; Length 105;</p> <p>Best Local Similarity 86.5%; Pred. No. 3e-09;</p> <p>Matches 90; Conservative 0; Mismatches 14; Indels 0; Gaps 0;</p>				
QY 25018	ATCCCAAGTCTTTGGGAGGCTAGCGGCTGGATCAGAGGTCTCAGGAGTTCAGAGCAGC	25077		
Db 105	ATCCCAAGTCTTTGGGAGGCTAGCGGCTGGATCAGAGGTCTCAGGAGTTCAGAGCAGC	46		
QY 25078	CTCCCAAGTCTTTGGGAGGCTAGCGGCTGGATCAGAGGTCTCAGGAGTTCAGGAGTTCAGGAGCAGC	25121		
Db 45	CTGCGCAATATGGTGAATCCCGTCTCTACTATAAGATATAAAAA	2		
<p>RESULT 10</p> <p>US-08-787-739-65/c</p> <p>Sequence 65, Application US/08787739</p> <p>Patent No. 6027887</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Zavada, Jan</p> <p>APPLICANT: Pastorekova, Silvia</p> <p>APPLICANT: Pastorek, Jaromir</p> <p>TITLE OF INVENTION: MN Gene and Protein</p> <p>NUMBER OF SEQUENCES: 96</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: Leona L. Lauder</p> <p>STREET: 369 Pine Street, Suite 610</p> <p>CITY: San Francisco</p> <p>STATE: California</p> <p>COUNTRY: USA</p> <p>ZIP: 94104</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Floppy disk</p> <p>COMPUTER: IBM PC compatible</p> <p>OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/08/787,739</p> <p>FILING DATE: 24-JAN-1997</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/485,049</p> <p>FILING DATE: 07-JUN-1995</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/486,756</p> <p>FILING DATE: 07-JUN-1995</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/477,504</p> <p>FILING DATE: 07-JUN-1995</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/481,658</p> <p>FILING DATE: 07-JUN-1995</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/485,862</p> <p>FILING DATE: 07-JUN-1995</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/485,863</p> <p>FILING DATE: 07-JUN-1995</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/487,077</p> <p>FILING DATE: 07-JUN-1995</p> <p>ATTORNEY/AGENT INFORMATION:</p> <p>NAME: Lauder, Leona L.</p> <p>REGISTRATION NUMBER: 30,863</p> <p>REFERENCE/DOCKET NUMBER: D-0021.4</p> <p>TELECOMMUNICATION INFORMATION:</p> <p>TELEPHONE: 415-981-2034</p> <p>TELEFAX: 415-981-0332</p> <p>INFORMATION FOR SEQ ID NO: 65:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 105 base pairs</p> <p>TYPE: nucleic acid</p> <p>STRANDEDNESS: double</p> <p>TOPOLOGY: linear</p> <p>MOLECULE TYPE: DNA (genomic)</p> <p>HYPOTHETICAL: NO</p> <p>ANTI-SENSE: NO</p> <p>US-08-787-739-65</p>				

Query Match 0.3%; Score 81.6; DB 5; Length 105;  
Best Local Similarity 86.5%; Pred. No. 3e-09;  
Matches 90; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
QY 25018 ATCCAGCAGCTTTGGAGGCGCTAGCGGTGATCAGGAGTTCAGACAGC 25077  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 105 ATCCAGCAGCTTTGGAGGCGCGAGCGTGGTATCACAAGTTCAGGAGTTTGAGAGCAGC 46  
QY 25078 CTCGCCAAGATGGTGAATCCGCTCTCTACTAAAGTATATAAA 25121  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 45 CTGGCAATATGGTGAACCGCTGTCTCTACTAAAGATGTAATAA 2

RESULT 11  
US-08-454-557C-91  
; Sequence 91, Application US/08454557C  
; Patent No. 5830670  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; APPLICANT: Wands, Jack R.  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; TITLE OF INVENTION: of Alzheimer's Disease  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, Steven R.  
; REGISTRATION NUMBER: 36,203  
; REFERENCE/DOCKET NUMBER: 0609.3840003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 91:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 84 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
US-08-454-557C-91

Query Match 0.2%; Score 67.2; DB 3; Length 84;  
Best Local Similarity 90.0%; Pred. No. 3.8e-06;  
Matches 72; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 11636 TGTAAACAGGATGCTCGATCTCTGACCTCGTGATCGGCCGCCCTCAGCCTCCCAA 11695  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 4 TGTTCATCAGGCTGGTGTGCAACTCTGACCTCGTGATCGGCCGCCCTCAGCCTCCCAA 63  
QY 11696 GTGCTGGGATTACAGGAGTG 11715  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 64 GTGCTGGGATTACAGCGTG 83

RESULT 12  
US-08-340-426D-91  
; Sequence 91, Application US/08340426D  
; Patent No. 5948634

GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; APPLICANT: Wands, Jack R.  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; TITLE OF INVENTION: of Alzheimer's Disease  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/340.426D  
; FILING DATE: 14-NOV-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, Steven R.  
; REGISTRATION NUMBER: 36,203  
; REFERENCE/DOCKET NUMBER: 0609.3840002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 91:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 84 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
US-08-340-426D-91

Query Match 0.2%; Score 67.2; DB 4; Length 84;  
Best Local Similarity 90.0%; Pred. No. 3.8e-06;  
Matches 72; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 11636 TGTAAACAGGATGCTCGATCTCTGACCTCGTGATCGGCCGCCCTCAGCCTCCCAA 11695  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 4 TGTTCATCAGGCTGGTGTGCAACTCTGACCTCGTGATCGGCCGCCCTCAGCCTCCCAA 63  
QY 11696 GTGCTGGGATTACAGGAGTG 11715  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 64 GTGCTGGGATTACAGCGTG 83

RESULT 13  
US-08-450-673C-91  
; Sequence 91, Application US/08450673C  
; Patent No. 5948888  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; APPLICANT: Wands, Jack R.  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; TITLE OF INVENTION: of Alzheimer's Disease  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450.673C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
;
US-08-450-673C-91

Query Match 0.2%; Score 67.2; DB 4; Length 84;
Best Local Similarity 90.0%; Pred. No. 3.8e-06;
Matches 72; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11636 TGTAAACAGGATGCTCGATCTCTGACCTCGTGATCGCGCCGCTCAGCTCCCAA 11695
Db 4 TGTTCATCAGGCTGGTGTGGAACCTCTGACCTCGTGATCGCGCCGCTCAGCTCCCAA 63

QY 11696 GTGCTGGGATTACAGAGTG 11715
Db 64 GTGCTGGGATTACAGCGTG 83

RESULT 14
PCT-US95-17111A-91
; Sequence 91, Application PC/TUS9517111A
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and
; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
;
US-08-454-557C-91

Query Match 0.2%; Score 62.2; DB 3; Length 84;
Best Local Similarity 84.3%; Pred. No. 4.9e-05;
Matches 70; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 25008 CAGCGCTATAATCCAGAGCTCTTTGGGAGGCGCTAGCGGGTGGATCAGAGGTCAGAGTT 25067
Db 83 CACGCTTGTAATCCAGACACTTTGGGAGGCTAGCGGGGCGGATCAGAGGTCAGAGTT 24

QY 25068 CAAGCCAGCGCTCGCCAAAGATGG 25090
Db 23 CGACACCGCGCTGATGAACATGG 1
```

```

;
; TOPOLOGY: both
;
PCT-US95-17111A-91

Query Match 0.2%; Score 67.2; DB 6; Length 84;
Best Local Similarity 90.0%; Pred. No. 3.8e-06;
Matches 72; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11636 TGTAAACAGGATGCTCGATCTCTGACCTCGTGATCGCGCCGCTCAGCTCCCAA 11695
Db 4 TGTTCATCAGGCTGGTGTGGAACCTCTGACCTCGTGATCGCGCCGCTCAGCTCCCAA 63

QY 11696 GTGCTGGGATTACAGAGTG 11715
Db 64 GTGCTGGGATTACAGCGTG 83

RESULT 15
US-08-454-557C-91/c
; Sequence 91, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,557C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
;
US-08-454-557C-91

Query Match 0.2%; Score 62.2; DB 3; Length 84;
Best Local Similarity 84.3%; Pred. No. 4.9e-05;
Matches 70; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 25008 CAGCGCTATAATCCAGAGCTCTTTGGGAGGCGCTAGCGGGTGGATCAGAGGTCAGAGTT 25067
Db 83 CACGCTTGTAATCCAGACACTTTGGGAGGCTAGCGGGGCGGATCAGAGGTCAGAGTT 24

QY 25068 CAAGCCAGCGCTCGCCAAAGATGG 25090
Db 23 CGACACCGCGCTGATGAACATGG 1
```

Search completed: June 15, 2000, 13:18:30  
Job time: 87840 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2000, 13:12:08 ; Search time 17972.2 Seconds  
(without alignments)  
-1569.757 Million cell updates/sec

Title: US-08-852-495C-l\_COPY\_84000\_113000  
Perfect score: 29001  
Sequence: 1 TTGTATTTTATTAGAGACA.....GCCTCAGCCCTCAGAGTGTA 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 370290

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vl:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*  
44: gb\_htg6:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hums:\*  
50: gb\_pl3:\*  
51: gb\_pr5:\*  
52: gb\_htg8:\*  
53: gb\_htg9:\*  
54: gb\_htg10:\*  
55: gb\_htg11:\*  
56: gb\_htg12:\*  
57: gb\_htg13:\*  
58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	95	0.3	107	9	HUMALCE162	M87924 Human carci
2	89	0.3	108	11	HSU67803	U67803 Human small
3	87.4	0.3	108	10	HSLDLRN2	X05250 Human LDL-r
4	86.8	0.3	108	10	HSLDLRN2	X05250 Human LDL-r
5	85.4	0.3	107	9	HUMALCE162	M87924 Human carci
6	83	0.3	103	9	HUMALCE221	M87896 Human carci
7	81.4	0.3	103	9	HUMALCE221	M87896 Human carci
8	81	0.3	108	11	HSU67803	U67803 Human small
9	80	0.3	108	10	HSLDLRD1	X05249 Human LDL-r
10	80	0.3	108	10	HSLDLRD2	X05251 Human LDL-r
11	79.4	0.3	108	10	HSLDLRD1	X05249 Human LDL-r
12	79.4	0.3	108	10	HSLDLRD2	X05251 Human LDL-r
13	78.2	0.3	108	10	HSLDLI12	X05248 Human LDL-r
14	77.8	0.3	108	11	HSU67804	U67804 Human small
15	76.4	0.3	108	11	HSU67808	U67808 Human small
16	76.4	0.3	110	9	HUMALCE43	M87900 Human carci
17	76	0.3	104	9	HUMALCE272	M87899 Human carci
18	75.4	0.3	97	9	HUMDLIRA2	M1480 Human low d
19	74	0.3	110	9	HUMALCE43	M87900 Human carci
20	73	0.3	108	11	HSU67808	U67808 Human small
21	72.6	0.3	106	13	G32743	G32743 A009P31 Hum
22	72.8	0.3	110	11	HSU67807	U67807 Human small
23	72.2	0.2	108	9	HUMD1D03M5	D16965 Human HepG2
24	72.4	0.2	108	10	HSLDLI12	X05248 Human LDL-r
25	72	0.2	90	9	HUMDLRFL	K03555 Human low d
26	71.2	0.2	103	13	HS8IC8R	X57789 Human sequ
27	71.4	0.2	108	11	HSU67804	U67804 Human small
28	70.8	0.2	91	13	HUMUT8164A	L30244 Human STS U
29	70.4	0.2	104	9	HUMALCE272	M87899 Human carci
30	70	0.2	107	11	HSU67806	U67806 Human small
31	69.4	0.2	110	11	HSU67807	U67807 Human small
32	68.8	0.2	80	9	HUMBRKFAE	M36135 Human alpha
33	68.2	0.2	107	11	HSU67806	U67806 Human small
34	67.2	0.2	80	9	HUMBRKFAE	M36135 Human alpha
35	67.4	0.2	94	9	HUMHGAL	M13479 Human alpha
36	67.4	0.2	95	13	HUMUT8002B	L30176 Human STS U
37	66.6	0.2	91	13	HUMUT8164A	L30244 Human STS U
38	66.8	0.2	94	9	HUMHGAL	M13479 Human alpha
39	66.8	0.2	95	13	HUMUT8002B	L30176 Human STS U
40	67	0.2	103	13	HS8IC8R	X57789 Human sequ
41	66.2	0.2	108	13	G43555	G43555 WIATF-2393-S
42	65.6	0.2	97	9	HOMDLDRDJ	M14179 Human famil
43	65.2	0.2	79	10	S73203	S73203 ALL-1 {tand
44	65.2	0.2	108	9	HUMD1D03M5	D16965 Human HepG2
45	64.8	0.2	76	11	AF032287	AF032287 Eulemur m

ALIGNMENTS

```

RESULT 1
HUMALCE162 107 bp ss-RNA PRI 15-APR-1994
LOCUS Human carcinoma cell-derived Alu RNA transcript, clone CE162.
DEFINITION M87924
ACCESSION M87924.1 GI:174871
VERSION 1
KEYWORDS Alu repeat.
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 107)
AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
post-transcriptional selection of master sequences
JOURNAL J. Mol. Biol. (1992) In press
FEATURES
    Location/Qualifiers
        1..107
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /cell_line="NTera2D1"
            /dev_stage="embryo"
            /sex="male"
            /tissue_type="carcinoma"
BASE COUNT 28 a 30 c 35 g 14 t
ORIGIN

Query Match 0.3%; Score 95; DB 9; Length 107;
Best Local Similarity 95.1%; Pred. No. 1e-08;
Matches 98; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 7316 GCACGAGAAATGGCTGACCGGGAGCGGAGCTTGACGTGAGCGGAGATCGGCCCATG 7375
|||||
Db 5 GGCAGAAATGGCTGACCGGGAGCGGAGCTTGACGTGAGCGGAGATCGGCCCATG 64

Qy 7376 GCACCTCCAGCTGGGCGACGAGCGAGACTCGTCTCAAAAAA 7418
|||||
Db 65 GCACCTCCAGCTGGGCGACGAGCGAGACTCGTCTCAAAAAA 107

RESULT 2
HSU67803 108 bp RNA PRI 01-AUG-1997
LOCUS Human small cytoplasmic Alu transcript.
DEFINITION U67803
ACCESSION U67803.1 GI:2289917
VERSION 1
KEYWORDS Alu.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE cDNAs derived from primary and small cytoplasmic Alu (scAlu)
transcripts
JOURNAL J. Mol. Biol. 271 (2), 222-234 (1997)
MEDLINE 97415756
REFERENCE 2 (bases 1 to 108)
AUTHORS Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The
Children's Hospital of Philadelphia, 1004F Abramson Research
Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA
FEATURES
    Location/Qualifiers
        1..108
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="TscAlu2"
            /note="scAlu"
repeat_region 1..108
            /rpt_family="Alu"

```

```

BASE COUNT 23 a 39 c 30 g 16 t
ORIGIN

Query Match 0.3%; Score 89; DB 11; Length 108;
Best Local Similarity 94.8%; Pred. No. 1.5e-07;
Matches 92; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2770 GCCTGTAATCTAGACATTTGGGAGCGGAGCGGCGGATCACGAGGTCAGGAGATCGA 2829
|||||
Db 1 GCCTGTAATCTAGACATTTGGGAGCGGAGCGGCGGATCACGAGGTCAGGAGATCGA 60

Qy 2830 GACCATCTTGGTACACGGTGAAACCCCGTTCTTAC 2866
|||||
Db 61 GACCATCTTGGTACACGGTGAAACCCCGTTCTTAC 97

RESULT 3
HSLDLRN2/c 108 bp DNA PRI 20-MAY-1992
LOCUS Human LDL-receptor gene intron 14 fragment (normal gene).
DEFINITION X05250
ACCESSION X05250.1 GI:34337
VERSION 1
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
Williamson,R. and Humphries,S.
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT See X05252 for deletion junction
FEATURES
    Location/Qualifiers
        1..108
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /note="intron XIV fragment"
BASE COUNT 28 a 23 c 39 g 18 t
ORIGIN

Query Match 0.3%; Score 87.4; DB 10; Length 108;
Best Local Similarity 89.5%; Pred. No. 3.1e-07;
Matches 94; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 9676 CTCGGCTCAGCGCAACCTCCGCTCCAGGGTTCAAGCAATTCCTCGCTCAGCTCCCG 9735
|||||
Db 108 CTCGGCTCAGTCAACCTCTGCTCTCGGTTCAAGCAATTCCTCGCTCAGCTCCCG 49

Qy 9736 AGTAATGGGACTACTGGCAAGCGCCACGCGCTGGCTAATTTT 9780
|||||
Db 48 AGTAGTGGGATTACAGGCACTGCCACACGCTGGCTAATTTT 4

RESULT 4
HSLDLRN2 108 bp DNA PRI 20-MAY-1992
LOCUS Human LDL-receptor gene intron 14 fragment (normal gene).
DEFINITION X05250
ACCESSION X05250.1 GI:34337
VERSION 1
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.

```



REFERENCE	1 (bases 1 to 108)
AUTHORS	Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R., Williamson,R. and Humphries,S.
TITLE	Unequal crossing-over between two alu-repetitive DNA sequences in the low-density-lipoprotein-receptor gene. A possible mechanism for the defect in a patient with familial hypercholesterolaemia
JOURNAL	Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE	87161901
COMMENT	See X05252 for deletion junction
FEATURES	Data kindly reviewed (07-DEC-1987) by HUMPHRIES S. Location/Qualifiers
source	1..108
organism	/organism="Homo sapiens"
db_xref	/db_xref="taxon:9606"
intron	1..108
note	/note="intron XIV fragment"
BASE COUNT	28 a 23 c 39 g 18 t
ORIGIN	
Query Match	0.3%; Score 86.8; DB 10; Length 108;
Best Local Similarity	88.7%; Pred. No. 4.1e-07;
Matches	94; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY	2877 AAAAATTAGCGGCGCTGTGGCGGGCCCTGTAGTCCCAGCTACTTTGGGAGGCTGAGGC 2936
Db	3 AAAAAATAGCAGCGGTGGTGCGAGGTGCCCTGTAAATCCCAGCTACTCGGGAGGCTGAGGC 62
QY	2937 AGGAATGGCATGAACCTGGGAGCGGAGCTTCAGTGAGCCGAG 2982
Db	63 AGGAGATTGCTTAACCCAGGAGCGAGAGTTTCAGTGAGCCGAG 108
RESULT	5
Locus	HUMALCE162/c 107 bp ss-RNA PRI 15-APR-1994
DEFINITION	Human carcinoma cell-derived Alu RNA transcript, clone CE162.
ACCESSION	M87924
VERSION	M87924.1 GI:174871
SOURCE	Alu repeat.
ORGANISM	Homo sapiens male embryo carcinoma cDNA to other RNA. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 107)
AUTHORS	Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE	Alu RNA transcripts in human embryonal carcinoma cells. Model of post-transcriptional selection of master sequences
JOURNAL	J. Mol. Biol. (1992) In press
FEATURES	Location/Qualifiers
source	1..107
organism	/organism="Homo sapiens"
db_xref	/db_xref="taxon:9606"
cell_line	/cell_line="Ntera2D1"
dev_stage	/dev_stage="embryo"
sex	/sex="male"
tissue_type	/tissue_type="carcinoma"
BASE COUNT	28 a 30 c 35 g 14 t
ORIGIN	
Query Match	0.3%; Score 85.4; DB 9; Length 107;
Best Local Similarity	89.3%; Pred. No. 7.7e-07;
Matches	92; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY	8540 TTTTTTGAGACTGAGCCCTTGCTCTGTCAACCAGGCTGGAGTGCATAGCGCGATCTCGGC 8599
Db	107 TTTTTTGAGAGGAGTCTCGCTCTGTCCGACGCTGGAGTGCAGTGGCGGATCTCGGC 48
QY	8600 TCATGTCAACCTCCGTTCCGAGTTCAAGCGATTCTACTGCC 8642
Db	47 TCATGTGAAGCTCCGCTCCCGGGTTTCACGCAATCTTCTCTGCC 5

```

RESULT 6
HUMALCE221      103 bp ss-RNA      PRI      15-APR-1994
LOCUS           Human carcinoma cell-derived Alu RNA transcript, clone CE221.
DEFINITION
ACCESSION      M87896
VERSION        M87896.1 GI:174874
KEYWORDS       Alu repeat.
SOURCE         Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 103)
AUTHORS       Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE         Alu RNA transcripts in human embryonal carcinoma cells. Model of
                post-transcriptional selection of master sequences
JOURNAL        J. Mol. Biol. (1992) In press
FEATURES       Location/Qualifiers
                1..103
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /cell_line="Ntera2D1"
                /dev_stage="embryo"
                /sex="male"
                /tissue_type="carcinoma"
BASE COUNT    25 a      27 c      33 g      18 t
ORIGIN

Query Match      0.3%; Score 83; DB 9; Length 103;
Best Local Similarity 89.9%; Pred. No. 2.3e-06;
Matches 89; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 7286 GCCTGTATGCCAGCTACTCGGGAGGCTGAGCGAGGAGATGGCGTGAACCGGGAGCG 7345
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        5 GCCTGTATGCCAGCTACTACACGGGAGCTAAGCGAGGAGATCGCTTGAACCGGGAGCG 64

QY 7346 GAGCTTCGAGTGAGCGAGATCGCGCATGCGCACTCCAG 7384
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        65 GAGGTGCGATGAGCGGAGATCGTGCCATTGCACTCCAG 103

RESULT 7
HUMALCE221/c   HUMALCE221      103 bp ss-RNA      PRI      15-APR-1994
LOCUS           Human carcinoma cell-derived Alu RNA transcript, clone CE221.
DEFINITION
ACCESSION      M87896
VERSION        M87896.1 GI:174874
KEYWORDS       Alu repeat.
SOURCE         Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 103)
AUTHORS       Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE         Alu RNA transcripts in human embryonal carcinoma cells. Model of
                post-transcriptional selection of master sequences
JOURNAL        J. Mol. Biol. (1992) In press
FEATURES       Location/Qualifiers
                1..103
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /cell_line="Ntera2D1"
                /dev_stage="embryo"
                /sex="male"
                /tissue_type="carcinoma"
BASE COUNT    25 a      27 c      33 g      18 t
ORIGIN

Query Match      0.3%; Score 81.4; DB 9; Length 103;
Best Local Similarity 88.9%; Pred. No. 4.7e-06;
Matches 88; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```



Query Match 0.3%; Score 80; DB 10; Length 108;  
 Best Local Similarity 85.6%; Pred. No. 8.8e-06;  
 Matches 89; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 9677 TCGGCTACCCGACACCTCCGCTCCAGGTTCAAGCAATTCCTCGCTCAGCTCCGCCA 9736

Db 107 TCGGCTACCCACACCTCGCTCCCTGGGTTCAACACCAATTTCTCGCTCAGCTCCGCCA 48

Qy 9737 GTAATTGGGACTACTGGCAAGCGCACCCAGCTGGCTAATTTT 9780

Db 47 GTAGCTGGGATTACAGGCACCTGCCACCACCGCTGGCTAAATTTT 4

RESULT 11  
 HSLDLR1/c  
 LOCUS Human LDL-receptor mutated gene with intron 12 deletion junction.  
 DEFINITION X05249  
 VERSION X05249.1 GI:34335  
 KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 108)

TITLE  
 Horsthemke, B., Beisiegel, U., Dunning, A., Havinga, J.R.,  
 Williamson, R., and Humphries, S.  
 Unequal crossing-over between two alu-repetitive DNA sequences in  
 the low-density-lipoprotein-receptor gene. A possible mechanism for  
 the defect in a patient with familial hypercholesterolaemia  
 Eur. J. Biochem. 164 (1), 77-81 (1987)

JOURNAL  
 MEDLINE  
 COMMENT  
 \*source: hypercholesterol aemia  
 See X05248 for corresponding normal gene sequence  
 in the defective LDL-receptor gene the deletion occurred between two  
 alu-repetitive sequences, that are in the same direction, the  
 deletion eliminates exons 13 and 14 and changes the reading frame  
 of the resulting spliced mRNA.  
 Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES  
 source  
 1. .108  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_type="blood leukocytes from a patient with familial"  
 misc\_feature 1. .108  
 /note="deletion junction region intron 12/ intron 15"  
 BASE COUNT 20 a 40 c 20 g 28 t  
 ORIGIN

Query Match 0.3%; Score 79.4; DB 10; Length 108;  
 Best Local Similarity 84.8%; Pred. No. 1.1e-05;  
 Matches 89; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 2877 AAAAATTAGCCGGCGTGTGGCGGCGCTGTACTGCCAGTCTTGGAGGCTGAGGC 2936

Db 106 AAAATTAGCCAGGCGTGTGGCAGGTGCTGTATCCAGCTACTCGGAGGCTGAGGC 47

Qy 2937 AGGAGAATGGCATGAACCTGGAGCGGAGCTTGTCAGTGAGCCGA 2981

Db 46 AGGAAATGGTTTGAACCCAGGAGGAGGTTGTGGTGAGCGCA 2

RESULT 12  
 HSLDLR2  
 LOCUS Human LDL-receptor mutated gene with intron 14 deletion junction.  
 DEFINITION X05251  
 ACCESSION X05251  
 VERSION X05251.1 GI:34336  
 KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

REFERENCE  
 AUTHORS Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 108)  
 Horsthemke, B., Beisiegel, U., Dunning, A., Havinga, J.R.,  
 Williamson, R., and Humphries, S.

TITLE  
 Unequal crossing-over between two alu-repetitive DNA sequences in  
 the low-density-lipoprotein-receptor gene. A possible mechanism for  
 the defect in a patient with familial hypercholesterolaemia  
 Eur. J. Biochem. 164 (1), 77-81 (1987)

JOURNAL  
 MEDLINE  
 COMMENT  
 \*source: hypercholesterol aemia

See X05250 for corresponding normal gene sequence  
 in the defective LDL-receptor gene the deletion occurred between two  
 alu-repetitive sequences, that are in the same direction, the  
 deletion eliminates exons 13 and 14 and changes the reading frame  
 of the resulting spliced mRNA.  
 Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES  
 source  
 1. .108  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_type="blood leukocytes from a patient with familial"  
 intron 1. .108  
 /note="intron XIV fragment"  
 BASE COUNT 28 a 20 c 40 g 20 t  
 ORIGIN

Query Match 0.3%; Score 79.4; DB 10; Length 108;  
 Best Local Similarity 84.8%; Pred. No. 1.1e-05;  
 Matches 89; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 2877 AAAAATTAGCCGGCGTGTGGCGGCGCTGTACTGCCAGTCTTGGAGGCTGAGGC 2936

Db 3 AAAAATTAGCCAGGCGTGTGGCAGGTGCTGTATCCAGCTACTCGGAGGCTGAGGC 62

Qy 2937 AGGAGAATGGCATGAACCTGGAGCGGAGCTTGTCAGTGAGCCGA 2981

Db 63 AGGAAATGGTTTGAACCCAGGAGGAGGTTGTGGTGAGCGCA 107

RESULT 13  
 HSLDL12  
 LOCUS Human LDL-receptor gene intron 12 fragment (normal gene) LDL - low  
 DEFINITION density lipoprotein.  
 ACCESSION X05248  
 VERSION X05248.1 GI:34334  
 KEYWORDS Alu repetitive sequence; low density lipoprotein receptor;  
 repetitive sequence.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 108)

REFERENCE  
 AUTHORS Horsthemke, B., Beisiegel, U., Dunning, A., Havinga, J.R.,  
 Williamson, R., and Humphries, S.  
 Unequal crossing-over between two alu-repetitive DNA sequences in  
 the low-density-lipoprotein-receptor gene. A possible mechanism for  
 the defect in a patient with familial hypercholesterolaemia  
 Eur. J. Biochem. 164 (1), 77-81 (1987)

JOURNAL  
 MEDLINE  
 COMMENT  
 \*source: hypercholesterol aemia

See X05249 for deletion junction  
 Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

FEATURES  
 source  
 1. .108  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 complement(<1. .65)  
 /note="Alu repeat"  
 intron 1. .108  
 /note="intron XII fragment"  
 BASE COUNT 21 a 38 c 20 g 29 t  
 ORIGIN

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE cDNAs derived from primary and small cytoplasmic Alu (sAlu)
transcripts
J. Mol. Biol. 271 (2), 222-234 (1997)
JOURNAL
MEDLINE 97415756
REFERENCE
AUTHORS Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The
Children's Hospital of Philadelphia, 1004F Abramson Research
Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA
FEATURES
source
1..108
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TscAlu7"
repeat_region
1..108
/note="sAlu"
/rpt_family="Alu"
/rpt_type="dispersed"
BASE COUNT 22 a 37 c 28 g 21 t
ORIGIN
Query Match 0.3%; Score 76.4; DB 11; Length 108;
Best Local Similarity 88.3%; Pred. No. 4.4e-05;
Matches 83; Conservative 0; Mismatches 11; Indels 0; Gaps
QY 14 AGAGACAGGGTTTCACATATGTTGGCCAGGCTGATCTCAAACTCCGTGACCTCATGATCCGC 73
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 95 AAAGACGGAGTTTCACCATATGTTGGCCAGGCTGGTCTCAAACTCCGTGACCTCATGATCCAC 36
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 74 CTGCCTGGGCCTCTCAAAGTGTCTGGGATTACAGG 107
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 35 CCGACITGGGCCTCCCAAGTGTCTGGGATTACAGG 2
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Search completed: June 16, 2000, 00:11:20
Job time: 127964 sec

```

Search completed: June 16, 2000, 00:11:20  
Job time: 127964 sec

Query Match	0.3%	Score 78.2;	DB 10;	Length 108;
Best Local Similarity	83.2%	Pred. No. 2e-05;		
Matches	89;	Conservative	0;	Mismatches 18; Indels 0; Gaps 0;

QY	9677	TCGGCTCACGCGAACCTCCGGCTCCAGGGTTCAAGCAATTCCTCTGCTCAGCCTCCGCCA	9736
DB	2	TCGGCTCACACACCTCTGCCTCTGGTTCAACCATTTTCTGCTCAGCCTCCTTA	61

QY	9737	GTAAATGGGACTACTGCAAGCGCCACACCGCTGGGTAAATTTGTA	9783
DB	62	GTAGCTGGGATTACAAGCATGTGCCACGCGCGGCTGATTTGTA	108

RESULT	14
HSU67804	
LOCUS	HSU67804 108 bp RNA PRI 01-AUG-1997
DEFINITION	Human small cytoplasmic Alu transcript.
ACCESSION	U67804
VERSION	U67804.1 GI:2289918
KEYWORDS	Alu.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE	(bases 1 to 108)
JOURNAL	Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The Children's Hospital of Philadelphia, 1004F Abramson Research Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA
FEATURES	Location/Qualifiers
source	1. .108
repeat_region	/organism="Homo sapiens" /db_xref="taxon.9606" /clone="fscAlu3" 1. .108 /note="scAlu" /rpt_family="Alu" /rpt_type=dispersed
BASE COUNT	26 a 38 c 26 g 18 t
ORIGIN	

Query Match	0.3%	Score 77.8;	DB 11;	Length 108;
Best Local Similarity	87.6%	Pred. No. 2.4e-05;		
Matches	85;	Conservative	0;	Mismatches 12; Indels 0; Gaps 0;

QY	2770	GCCTGTAATCCTAGCACCTTTGGAGGCGGAGACGGGGGATCACAGGTACGAGATCGA	2829
DB	1	GCCTGTAATCCACGACTTTGGAAAGGCGAAAGCGGGAGGATCAAGGTACGAGATCGA	60

QY	2830	GACCATCTTGCTAACACGCTGAACCCCGCTTCTAC	2866
DB	61	GACCATCTTGCTAACACGCTGAACCCCGCTTCTAC	97

RESULT	15
HSU67808/c	
LOCUS	HSU67808 108 bp RNA PRI 01-AUG-1997
DEFINITION	Human small cytoplasmic Alu transcript.
ACCESSION	U67808
VERSION	U67808.1 GI:2289922
KEYWORDS	Alu.
SOURCE	human.
ORGANISM	Homo sapiens

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2000, 16:12:36 ; Search time 593.87 seconds  
(without alignments)  
12217.860 Million cell updates/sec

Title: US-08-852-495C-1\_COPY\_84000\_113000  
Perfect score: 29001  
Sequence: 1 TTGTAATTTTATTAGACAC.....GCCTCAGCCTTCAGAGTGA 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 433070

Minimum DB seq length: 10

Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	67.2	0.2	100	1 T24892	Human gene signatu
C 2	67	0.2	108	1 T26828	Human gene signatu
C 3	66.8	0.2	108	1 X12095	Human biallelic po
C 4	65.6	0.2	100	1 T24892	Human gene signatu
C 5	63.8	0.2	86	1 V41231	Mouse embryonic ce
C 6	62.8	0.2	103	1 T20927	Human gene signatu
C 7	61	0.2	108	1 T25009	Human gene signatu
C 8	60	0.2	93	1 T25688	Human gene signatu
C 9	58.6	0.2	108	1 T25009	Human gene signatu
C 10	58.6	0.2	110	1 T26288	Human gene signatu
C 11	58.4	0.2	110	1 T25260	Human gene signatu
C 12	57	0.2	108	1 T26828	Human gene signatu
C 13	57.2	0.2	108	1 X12095	Human biallelic po
C 14	56.2	0.2	103	1 T26213	Human gene signatu
C 15	54.8	0.2	97	1 T26728	Human gene signatu
C 16	55	0.2	100	1 X12087	Human biallelic po
C 17	55	0.2	100	1 X12085	Human biallelic po
C 18	55	0.2	100	1 X12086	Human biallelic po
C 19	54.6	0.2	69	1 Q29016	Probe to internal
C 20	54.6	0.2	91	1 T25854	Human gene signatu
C 21	54.4	0.2	109	1 T23895	Human gene signatu
C 22	53.8	0.2	99	1 T20931	Human gene signatu
C 23	53.4	0.2	97	1 T26728	Human gene signatu
C 24	53.6	0.2	100	1 Q76490	Human genome fragm
C 25	53	0.2	82	1 T25468	Human gene signatu
C 26	52	0.2	101	1 T24667	Human gene signatu
C 27	52.2	0.2	110	1 T26288	Human gene signatu
C 28	51.4	0.2	91	1 T25854	Human gene signatu
C 29	51.4	0.2	102	1 T20743	Human gene signatu
C 30	51	0.2	88	1 T21564	Human gene signatu
C 31	51	0.2	89	1 T23513	Human gene signatu
C 32	50.4	0.2	62	1 T25689	Human gene signatu
C 33	50.4	0.2	95	1 T23131	Human gene signatu
C 34	50.6	0.2	100	1 X12085	Human biallelic po

C 35	50.6	0.2	100	1 X12086	Human biallelic po
C 36	49.8	0.2	74	1 T25218	Human gene signatu
C 37	49.6	0.2	84	1 T25848	Human gene signatu
C 38	49.8	0.2	103	1 T26213	Human gene signatu
C 39	49.2	0.2	70	1 N60231	Normal chromosome
C 40	49.2	0.2	95	1 Q75099	Plasmid pOKSC18a c
C 41	49.4	0.2	100	1 X12087	Human biallelic po
C 42	48.4	0.2	89	1 T23513	Human gene signatu
C 43	48.4	0.2	93	1 T22572	Human gene signatu
C 44	48	0.2	48	1 Q69409	Human H4/a gene fo
C 45	48	0.2	48	1 T63871	Human H4/a gene (f

#### ALIGNMENTS

RESULT 1  
T24892/c  
ID T24892 standard; cDNA to mRNA; 100 BP.  
AC T24892;  
DT 05-NOV-1996 (first entry)  
DE Human gene signature HUMGS06998.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN W09514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
PI Matsubara K. Okubo K;  
DR WPI; 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1; Page 1720; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 100 BP; 28 A; 22 C; 25 G; 22 T;

Query Match 0.2%; Score 67.2; DB 1; Length 100;  
Best Local Similarity 78.8%; Pred. No. 0.0031;  
Matches 78; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 8535 TTTTGTGTTTGTGAGCTGAGCCCTGCTGTCTACCCAGGCTGAGTGCATGCGGCATC 8594

Db 100 TTTGTTGTTTCAACAGAGTGTACTCTGTCCACGCGGAGTGCATGCGGCATC 41

QY 8595 TCGGCTCACTGCAACCTCCGCTTCCAGGTTCAAGCGAT 8633

Db 40 TCAGCTNATGCAAAATCTGCCTCCCGAGTTCAAGCGAT 2

RESULT 2

T26828/c  
ID T26828 standard; cDNA to mRNA; 108 BP.

T26828;  
14-NOV-1996 (first entry)  
Human gene signature HUMGS09078.  
Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
human; cloning; mapping; non-biased library; diagnosis; detection;  
cell typing; abnormal cell function; ss.  
Homo sapiens.  
WO9514772-A1.  
01-JUN-1995.  
11-NOV-1994; J01916.  
12-NOV-1993; JP-355504.  
(MATS/) MATSUBARA K.  
(OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
WPI; 95-206931/27.  
DR Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1; Page 2182; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 108 BP; 18 A; 33 C; 23 G; 28 T;  
  
Query Match 0.2%; Score 67; DB 1; Length 108;  
Best Local Similarity 90.9%; Pred. No. 0.0034;  
Matches 70; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
Qy 2751 GCCGGCGGTGGTCAAGCTGTAATCTAGACTTTGGAGGCCGAGACGGGCGGAT 2810  
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 77 GCCGGCGGTGGTCAAGCTGTAATCTAGACTTTGGAGGCCGAGACGGGCGGAT 18  
  
Qy 2811 CACGAGGTCAGGATC 2827  
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 17 GACGAGGTCAGGAGATC 1  
  
RESULT 3  
X12095  
ID X12095 standard; DNA; 108 BP.  
AC X12095;  
DT 30-MAR-1999 (first entry)  
DE Human biallelic polymorphic DNA fragment TIGR-A003M18a.  
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;  
KW detection; phenotypic typing; characteristic; infection; hereditary;  
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;  
KW treatment; marker; ss.  
OS Homo sapiens.  
PN WO9820165-A2.  
PD 14-MAY-1998.  
PE 05-NOV-1997; U20313  
PR 06-NOV-1996; US-030455.  
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
PI Hudson T, Lander ES, Wang D;  
WPI; 98-286974/25.  
DR New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease  
PS Claim 1; Page 219; 310pp; English.

CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
CC markers which have been isolated using the primers represented in  
CC X09121-X10268. The base occupying the polymorphic site is indicated by  
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.  
SQ Sequence 108 BP; 19 A; 23 C; 28 G; 37 T;  
  
Query Match 0.2%; Score 66.8; DB 1; Length 108;  
Best Local Similarity 81.5%; Pred. No. 0.0037;  
Matches 88; Conservative 1; Mismatches 18; Indels 1; Gaps 1;  
  
Qy 9780 TGTATTTTATTAGAGATGAGGTTTCTCCATGTTGTCAGACTGCTCGAACTCCGAC 9839  
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 1 TGTCTTTTGTAGAGATGAGGTTTCTCTGTGGCCAGGATGCTCGAACTCCTGAC 60  
  
Qy 9840 CTCAGGTGATCACCGCTCGGCTCC-AAAAGTCTGGGATTACAG 9886  
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 61 TTCAAGTCATCGTCTGCTGGCTCCCAAAAGTCTGGGATTATAG 108  
  
RESULT 4  
T24892  
ID T24892 standard; cDNA to mRNA; 100 BP.  
AC T24892;  
DT 05-NOV-1996 (first entry)  
DE Human gene signature HUMGS06998.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN WO9514772-A1.  
PD 01-JUN-1995.  
PE 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
WPI; 95-206931/27.  
DR Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1; Page 1720; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.











```

PT tissues
PS Claim 1; Page 2158; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared from
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 97 BP; 19 A; 27 C; 20 G; 28 T;

Query Match          0.28; Score 54.8; DB 1; Length 97;
Best Local Similarity 73.9%; Pred. No. 0.41;
Matches            68; Conservative      0; Mismatches    24; Indels     0; Gaps     0;

QY 14573  GATCTGCCGCGCTGGCCTCCCAAAGTGTTCGGATTACAGGCATCAGGCAGTGCCTCCAC 14632
         ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       1   GATCTGCCCACTNGGCGCTCCCAAGTGCTCGGATTACAGCATCAGGCATCGCCCCG 60

QY 14633  CCAGGACAGATTTTTTTTACACTCATGTTTCT 14664
         | | | | | | | | | | | | | | | | | | | |
Db       61  NCTGACTAAGTCATCTTTTTTTTTTAATTTCT 92

```

Search completed: June 16, 2000, 00:47:51  
Job time: 128982 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2000, 08:12:30 ; Search time 8513.49 Seconds  
(without alignments)  
13807.208 Million cell updates/sec

Title: US-08-852-495C-1\_COPY\_84000\_113000  
Perfect score: 29001  
Sequence: 1 TTGTATTTTATTAGAGACA.....GCCTCAGCCTTCAGAGTGTA 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues  
Total number of hits satisfying chosen parameters: 156056

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : EST:  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result	%	Query
SUMMARIES		

No.	Score	Match	Length	DB	ID	Description
C 1	96.2	0.3	109	30	AA243009	AA243009 zr25h02.s
C 2	94.8	0.3	106	37	AA703692	AA703692 ag81a10.r
C 3	93.8	0.3	105	61	AA828832	AA828832 at72909.x
C 4	91.8	0.3	103	84	AA8914	AA8914 RPI11-4A12
C 5	91.6	0.3	106	38	AA812141	AA812141 ob48h02.s
C 6	91.2	0.3	107	39	AA828124	AA828124 od71a07.s
C 7	91.4	0.3	109	30	AA243009	AA243009 zr25h02.s
C 8	89.8	0.3	101	39	AA835205	AA835205 ak64h01.s
C 9	90	0.3	109	24	N25299	N25299 yw52c09.s1
C 10	89.6	0.3	97	25	AA9638	AA9638 yv25e09.r1
C 11	88.6	0.3	103	94	AA028649	AA028649 CIT-RSP-2
C 12	88.2	0.3	101	94	AA076649	AA076649 CIT-RSP-2
C 13	88.4	0.3	105	105	AA0264176	AA0264176 CITR-EI-
C 14	87.6	0.3	103	38	AA807640	AA807640 nx08h05.s
C 15	87.6	0.3	110	79	AA250394	AA250394 2822460.3
C 16	86.6	0.3	102	30	AA226656	AA226656 nc19f09.s
C 17	86.8	0.3	106	37	AA703692	AA703692 ag81a10.r
C 18	86.8	0.3	110	39	AA897366	AA897366 am06h02.s
C 19	86.4	0.3	107	33	AA385808	AA385808 EST99495
C 20	86	0.3	103	108	AA082186	AA082186 RPI1-11-4
C 21	85.8	0.3	106	108	AA044957	AA044957 CITR-EI-
C 22	85.8	0.3	107	24	H67040	H67040 yu68c01.r1
C 23	86	0.3	110	64	AA083640	AA083640 xc49f02.x
C 24	85.4	0.3	103	84	AA8914	AA8914 RPI11-4A12
C 25	85.2	0.3	107	35	AA565533	AA565533 nk42b11.s
C 26	85	0.3	109	84	BL7434	BL7434 345K2.IVB.C
C 27	85	0.3	110	94	AA003188	AA003188 RPI11-ID
C 28	83.8	0.3	103	30	AA228795	AA228795 nc14e07.s
C 29	83.8	0.3	105	109	AA0637292	AA0637292 RPI1-11-4
C 30	84	0.3	110	33	AA442529	AA442529 zv68h02.r
C 31	83.6	0.3	106	34	AA516339	AA516339 ng71a02.s
C 32	83.6	0.3	110	25	N72778	N72778 yv49a10.r1
C 33	83.4	0.3	110	94	AA003188	AA003188 RPI11-ID
C 34	82.8	0.3	102	36	AA654562	AA654562 nt75f10.s
C 35	82.8	0.3	105	74	AA196212	AA196212 xm06e06.x
C 36	82.4	0.3	101	35	AA583697	AA583697 nn58f10.s
C 37	82.6	0.3	108	32	AA370029	AA370029 EST81584
C 38	82	0.3	100	35	AA564832	AA564832 nj22a06.s
C 39	82.2	0.3	103	94	AA028649	AA028649 CIT-RSP-2
C 40	82.2	0.3	103	108	AA0384425	AA0384425 RPI1-11-4
C 41	82	0.3	106	94	AA062963	AA062963 CIT-RSP-2
C 42	82.2	0.3	106	105	AA0282340	AA0282340 RPI11-80
C 43	81.8	0.3	101	35	AA583697	AA583697 nn58f10.s
C 44	82	0.3	106	38	AA812141	AA812141 ob48h02.s
C 45	82	0.3	109	24	N25299	N25299 yw52c09.s1

## ALIGNMENTS

RESULT 1  
 AA243009/c  
 LOCUS  
 DEFINITION zr25h02.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens  
 cDNA clone IMAGE:664467 3' similar to contains Alu repetitive  
 element:contains element LTR1 repetitive element ;, mRNA sequence.

ACCESSION AA243009.1 GI:1873869  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 109)  
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
 Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
 Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
 WashU-NCI human EST Project  
 Unpublished (1997)  
 On Dec 3, 1996 this sequence version replaced gi:1126869.

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 1127 Std Error: 0.00  
 Seq primer: -4ml3 fwd. EF from Amersham  
 High quality sequence stop: 102.

## FEATURES

source  
 1. 109  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="GDB:5426481"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:664467"  
 /clone\_lib="Stratagene NT2 neuronal precursor 937230"  
 /tissue\_type="neuroepithelial cells"  
 /dev\_stage="Ntera-2 neuroepithelial cells"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: brain; Vector: pBluescript SK-; Site:1:  
 EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:  
 Oligo dt. Uninduced, exponentially growing neuroepithelial  
 cells (Ntera-2/ci.D1). Average insert size: 1.0 kb;  
 Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG  
 3' -3' adaptor sequence: 5' CTCGAGTGTGTTTTTTTTT 3'"

BASE COUNT 19 a 30 c 30 g 30 t  
 ORIGIN

Query Match 0.3% Score 96.2; DB 30; Length 109;  
 Best Local Similarity 92.7% Pred. No. 0.065; Indels 0; Gaps 0;  
 Matches 101; Conservative 0; Mismatches 8;  
 QY 2767 CACGCCTGTAATCTAGCATTGGAGCGCCGAGCAGCGGATCAGGAGTACAGAT 2826  
 Db 109 CACGCCTGTAATCTAGCATTGGAGCGCGAGTGGCGGATCAGGAGTACAGAT 50  
 QY 2827 CGACCATCTTGGCTACACGGTGAACCCGTTCTTCTACTAAAAATAC 2875  
 Db 49 CAAGACCATCTTGGCTACACGGTGAACCCGTTCTTCTACTAAAAATAC 1

RESULT 2  
 AA703692/c  
 LOCUS  
 DEFINITION ag81a10.r1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone  
 IMAGE:1140858 5' similar to contains Alu repetitive element; , mRNA  
 sequence.

ACCESSION AA703692.1 GI:2713610  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 106)  
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
 Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
 Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
 WashU-NCI human EST Project  
 Unpublished (1997)  
 On Sep 12, 1996 this sequence version replaced gi:1397630.

CONTACT: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.



Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 53.

#### FEATURES

source

1. .106  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1140858"  
/clone\_lib="Stratagene hNT neuron (#937233)"  
/dev\_stage="hNT neurons"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dr. Differntiated, post mitotic hNT neurons. Average insert size: 1.5 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3' "

19 a 29 c 29 g 29 t

#### BASE COUNT

ORIGIN

Query Match 0.3%; Score 94.8; DB 37; Length 106;  
Best Local Similarity 93.4%; Pred. No. 0.094; 7; Indels 0; Gaps 0;  
Matches 99; Conservative 0; Mismatches 0; Gaps 0;

QY 2767 CACGCTGTAATCTAGCACTTTGGAGCGCGGATCAGGATCAGGATCAGGAT 2826

Db 106 CACGCTGTAATCTAGCACTTTGGAGCGCGGATCAGGATCAGGATCAGGAT 47

QY 2827 CGAGCACCCTTGGCTAACACGGTGAACCCCGTTCTACTAAAAA 2872

Db 46 CGAGCACCCTTGGCTAACACGGTGAACCCCGTTCTACTAAAAA 1

#### RESULT 3

AI832832

LOCUS

AI832832 105 bp mRNA EST 13-JUL-1999  
at72g09.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone  
IMAGE:2377600 3' similar to contains Alu repetitive  
element; contains element MER22 repetitive element ;, mRNA sequence.

ACCESSION

AI832832

VERSION

AI832832.1 GI:5454812

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 105)

AUTHORS

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.

TITLE

WashU-NCI human EST Project

JOURNAL

Unpublished (1997)

COMMENT

On Dec 20, 1995 this sequence version replaced gi:1133644.  
Washington University School of Medicine  
Contact: Wilson RK  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40UP from Gibco.

#### FEATURES

source

1. .105  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2377600"  
/clone\_lib="Barstead colon HPLRB7"  
/sex="male"  
/dev\_stage="adult, age 25"  
/lab\_host="DH108 (phage resistant)"  
/note="Organ: colon; Vector: p7T3D-Pac (Pharmacia) with a

modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st  
strand cDNA was primed with a Not I - oligo(NT) primer (5'  
TGTTACGAATCTGAAGTGGAGCGGCCCTTTTCTTTTCTTTTCTTTTCTTTT  
3'); double-stranded cDNA was ligated to Eco RI adaptors  
[5' AATCACTAGTAAT 3' and 5' ATTACTAGTG 3'], digested  
with Not I and cloned into the Not I and Eco RI sites of  
the modified p7T3 vector. Library constructed by Bob  
Barstead."

17 a 35 c 27 g 26 t

ORIGIN

Query Match 0.3%; Score 93.8; DB 61; Length 105;  
Best Local Similarity 93.3%; Pred. No. 0.12;  
Matches 98; Conservative 0; Mismatches 0; Gaps 0;

QY 9627 GAGACGAGTTTCACACTTTGTTGCCCGAGTGGAGTGCATGTCGCTCACC 9686

Db 1 GAGACGAGTTTCGCTCTTTGTTGCCCGAGTGGAGTGCATGTCGCTCACC 60

QY 9687 GCAACCTCCGCTCCAGGCTTCAAGCAATTCCTCGCTCAGCCT 9731

Db 61 GCAACCTCCGCTCCGCTTCAAGCAATTCCTCGCTCAGCCT 105

#### RESULT 4

LOCUS

B48914

DEFINITION

RPC111-4A12.TP RPCI-11 Homo sapiens genomic clone

ACCESSION

B48914

VERSION

B48914.1 GI:2601151

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 103)

AUTHORS

Adams, M.D., Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.

TITLE

Use of BAC End Sequences for Sequence-Ready Map Building

JOURNAL

Unpublished (1997)

COMMENT

Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from  
Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/tdb/hungen/bac\_end\_search/bac\_end\_search.html  
Seq primer: SP6  
Class: BAC ends.

FEATURES

Location/Qualifiers

source

1. .103

/organism="Homo sapiens"

/db\_xref="GDB:7501163"

/db\_xref="taxon:9606"

/clone="RPCI-11-4A12"

/clone\_lib="RPCI-11"

/sex="Male"

/cell\_type="Lymphocytes"

/note="Vector: pBACE3 6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPC111 Human Male BAC Library"

30 a 28 c 30 g 15 t

BASE COUNT

ORIGIN

Query Match 0.3%; Score 91.8; DB 84; Length 103;  
Best Local Similarity 93.2%; Pred. No. 0.21;  
Matches 96; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2774 GTAATCTAGCACTTTGGAGCGCCGAGACGGCGGATCACGAGGTTCAGGATCGAGACC 2833  
|||||  
DB 1 GTAAGCCAGCACTTTGGAGCGCCGAGACGGCGGATCACGAGGTTCAGGATCGAGACC 60  
|||||

QY 2834 ATCTTGGCTATACACGGTGAACCCGGTCTCTACTAAAAATACA 2876  
|||||  
DB 61 ATCCCGGTAAACCGTGAACCCGGTCTCTACTAAAAATACA 103  
|||||

RESULT 5  
AA812141/C  
LOCUS AA812141 106 bp mRNA EST 19-FEB-1998  
DEFINITION ob48h02.s1 NCI-CGAP.GC81 Homo sapiens cDNA clone IMAGE:1334643 3'  
similar to contains Alu repetitive element;; mRNA sequence.  
ACCESSION AA812141  
VERSION AA812141.1 GI:2881752  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 106)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1402063.  
Contact: Robert Strausberg, Ph.D.  
Tel.: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
Ph.D., Gerald Marti, M.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1450 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 60.  
FEATURES  
source  
1..106  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1334643"  
/clone\_lib="NCI-CGAP\_GC81"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="vector: pT73P-Pac (Pharmacia) with a modified  
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),  
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
primed with a Not I - oligo(dT) primer  
[5'-TGTTACCAATCTGAAGTGGGCGGCGCTCATTTTTTTTTTTT-  
3']. Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

16 a 31 c 24 g 35 t  
BASE COUNT  
ORIGIN

Query Match 0.3%; Score 91.6; DB 38; Length 106;  
Best Local Similarity 92.3%; Pred. No. 0.24;  
Matches 96; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Best Local Similarity 91.5%; Pred. No. 0.21;  
Matches 97; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 7320 GGAGAATGGCTGAACGGGAGCGGAGCTTGAGTCAGCGGAGATCGCCGAC 7379  
|||||  
DB 106 GGAGAATGGCTGAACCTGGGAGGTGGAGCTTGAGTCAGCGGAGATCACACCTGCAC 47  
|||||

QY 7380 TCCAGCCTGGGTGACAGAGCGAGACTCCGTCTCAAAAAA 7425  
|||||  
DB 46 TCCAGCCTGGGTGACAGAGCGAGACTCCATCTCAAAAAA 1  
|||||

RESULT 6  
AA828124  
LOCUS AA828124 107 bp mRNA EST 20-FEB-1998  
DEFINITION od71a07.s1 NCI-CGAP\_Ov2 Homo sapiens cDNA clone IMAGE:1373364  
similar to contains Alu repetitive element; contains element MER22  
repetitive element ;, mRNA sequence.  
ACCESSION AA828124  
VERSION AA828124.1 GI:2900487  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 107)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 17, 1998 this sequence version replaced gi:1899815.  
Contact: Robert Strausberg, Ph.D.  
Tel.: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: David B. Krizman, Ph.D.  
DNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 93.  
FEATURES  
source  
1..107  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1373364"  
/clone\_lib="NCI-CGAP\_Ov2"  
/sex="female"  
/tissue\_type="ovary"  
/lab\_host="DH10B"  
/note="vector: pAMP10; mRNA made from invasive ovarian  
tumor, cDNA made by oligo-dT priming. Non-directionally  
cloned. Size-selected on agarose gel, average insert size  
600 bp. Reference: Krizman et al. (1996) Cancer Research  
56:5380-5383."

30 a 23 c 38 g 16 t  
BASE COUNT  
ORIGIN

Query Match 0.3%; Score 91.2; DB 39; Length 107;  
Best Local Similarity 92.3%; Pred. No. 0.24;  
Matches 96; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2862 TCTACTAAAAATACAAAAAATAGCGGCGCTGTGGCGGGCGCTGTAGTCCACGCTAC 2921  
|||||  
DB 4 TCGACTAAAAATACAAAAAATAGCCAGCGCTAATGGCGGCACCTGTAGTCCAGCTGC 63  
|||||

QY 2922 TTGGAGGCTGAGGAGGAGATGCGATGCACTGGGAGGGGA 2965  
|||||

```
Db 64 TTGGAGGCTGAGCAGGAGATGGCGTGAACCCGGGAGCGGA 107
RESULT 7
AA243009
LOCUS AA243009 109 bp mRNA EST 11-MAR-1998
DEFINITION cDNA clone IMAGE:664467 3' similar to contains Alu repetitive
element;contains element LTR1 repetitive element ;, mRNA sequence.
ACCESSION AA243009
VERSION AA243009.1 GI:1873869
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 109)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
TITLE Unpublished (1997)
JOURNAL
COMMENT On Dec 3, 1996 this sequence version replaced gi:1126869.
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1127 Std Error: 0.00
Seq Primer: 41m13 fwd. ET from Amersham
High quality sequence stop: 102.
Location/Qualifiers
1..109
/organism="Homo sapiens"
/db_xref="GDB:5426481"
/db_xref="taxon:9606"
/clone="IMAGE:664467"
/clone_lib="Stratagene NT2 neuronal precursor 937230"
/tissue_type="neuroepithelial cells"
/dev_stage="Ntera-2 neuroepithelial cells"
/lab_host="SOLR (kanamycin resistant)"
/Note="Organ: brain; Vector: pluescript SK-; Site:1:
EcoRI; Site:2: xhoI; Cloned unidirectionally. Primer:
Oligo dt. Uninduced, exponentially growing neuroepithelial
cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;
Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCAGG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3'
```

```
BASE COUNT 19 a 30 c 30 g 30 t
ORIGIN
Query Match 0.3%; Score 91.4; DB 30; Length 109;
Best Local Similarity 89.9%; Pred. No. 0 22;
Matches 98; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 8700 GTATTTTATGACAGAGGGTTTACCGTGTGGCCAGAGTTCCTCAATCTCCTTACC 8759
Db 1 GTATTTTATGACAGAGGGTTTACCGTGTGGCCAGAGTTCCTCAATCTCCTTACC 60

QY 8760 TCGTGATCGCCCGCTCTCTCTGCGCAAGTCTCGGATTACAGCGTG 8808
Db 61 TCGTGATCGCCCGCACCTCGCGCTCCCAAGTCTCGGATTACAGCGGTG 109

RESULT 8
AA835205/c
LOCUS AA835205 101 bp mRNA EST 23-FEB-1998
DEFINITION ak64h01.s1 Barstead pancreas HPLRB1 Homo sapiens cDNA clone
IMAGE:1412689 3' similar to contains Alu repetitive
```

```
element;contains element KER repetitive element ;, mRNA sequence.
ACCESSION AA835205
VERSION AA835205.1 GI:2908933
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 101)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
TITLE Unpublished (1997)
JOURNAL
COMMENT On Nov 29, 1993 this sequence version replaced gi:636191.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq Primer: -40m13 fwd. ET from Amersham.
Location/Qualifiers
1..101
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1412689"
/clone_lib="Barstead pancreas HPLRB1"
/sex="female"
/dev_stage="adult, 34 years"
/lab_host="DH10B"
/Note="Organ: pancreas; Vector: pTT73D-Pac (Pharmacla)
with a modified polylinker; Site_1: EcoRI; Site_2: NotI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5',
TGTTACGAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGGATCCTTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pTT73 vector.
Library constructed by Bob Barstead."
BASE COUNT 14 a 36 c 27 g 24 t
ORIGIN
Query Match 0.3%; Score 89.8; DB 39; Length 101;
Best Local Similarity 93.1%; Pred. No. 0.35;
Matches 94; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 7313 TGAGGCAGAGATGCGGTGAACCGCGGAGCGGAGCTTGCAGTGCAGCGATCGGCC 7372
Db 101 TGAGGCAGAGATGCGGTGAACCGCGGAGCGGAGCTTGCAGTGCAGCGATCAAGCC 42

QY 7373 ATGGCACTCCAGCTGGGTGACAGAGCGAGACTCCGCTCTCA 7413
Db 41 ACTGCACTCCAGCTGGCGGACAGAGTGCAGACTCCGCTCTCA 1

RESULT 9
N25299/c
LOCUS N25299 109 bp mRNA EST 28-DEC-1995
DEFINITION yw52c09.s1 Weizmann Olfactory Epithelium Homo sapiens cDNA clone
IMAGE:255856 3' similar to contains Alu repetitive element; , mRNA
sequence.
ACCESSION N25299
VERSION N25299.1 GI:1139449
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

REFERENCE  
AUTHORS

1 (bases 1 to 109)  
Hillier, L., Lennon, G., Becker, M., Bonaudo, M.F., Chiapelli, B.,  
Chisoe, S., Dietrich, N., DuBuque, T., Favell, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierly-Meg, J.,  
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
and Marra, M.

TITLE  
JOURNAL  
MEDLINE  
COMMENT

Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
9704478  
On Apr 14, 1993 this sequence version replaced gi:8377394.  
Contact: Wilton RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: mi3 -40 forward  
High quality sequence stop: 307.

FEATURES  
source

1. .109  
/organism="Homo sapiens"  
/db\_xref="GDB:3866265"  
/db\_xref="taxon:9606"  
/clone="IMAGE:255856"  
/clone\_lib="Weizmann Olfactory Epithelium"  
/sex="Female"  
/tissue\_type="olfactory epithelium"  
/dev\_stage="35 year old"  
/lab\_host="SOIL cells (kanamycin resistant)"  
/note="Organ: nose; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Olfactory epithelium, normal. Average insert  
size: 0.8 kb; Uni-ZAP XR Vector. Library constructed by N.  
Walker, D. Lancet, Weizmann Institute of Science. -5'  
adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor  
sequence: 5' CTCAGTTTCTTTTCTTTT 3"

BASE COUNT 13 a 34 c 24 g 35 t 3 others

ORIGIN

Query Match 0.3%; Score 90; DB 24; Length 109;  
Best Local Similarity 88.1%; Pred. No. 0.32;  
Matches 96; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 2933 AGGCAGGAGATGGCATGAACCTGGGAGGCGAGCTTGCAGTGCAGCGAGATTGGCCAC 2992  
|||||  
Db 109 AGGCAGGAGATGGCAGAACCTGGGAGGCGAGCTTGCAGTGCAGCGAGATCAGCCAC 50  
|||||

Qy 2993 TGCATCCCACTGGGAGACAGCGAGACTCCGCTCAAAAAA 3041  
|||||  
Db 49 TGCATCCAGCTGGGAGACAGCGAGAGTCCGCTCAAAAAA 1  
|||||

RESULT 10  
N49638  
LOCUS  
DEFINITION

N49638 97 bp mRNA EST 14-FEB-1996  
yv25e09.r1 Soares fetal liver spleen lNFLS Homo sapiens cDNA clone  
IMAGE:243784 5' similar to gb:X57138\_rnal HISTONE H2B.2 (HUMAN);,  
mRNA sequence.

ACCESSION N49638  
VERSION N49638.1 GI:1190804  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 97)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)  
On Apr 14, 1993 this sequence version replaced gi:693230.  
Contact: Wilton RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Trace considered overall poor quality  
Seq primer: T7  
High quality sequence stop: 1.

FEATURES  
source

1. .97  
/organism="Homo sapiens"  
/db\_xref="GDB:3792917"  
/db\_xref="taxon:9606"  
/clone="IMAGE:243784"  
/clone\_lib="Soares fetal liver spleen lNFLS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer  
[5' AACTGGAAGAAATAATTAAGATCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 24 a 29 c 23 g 20 t 1 others

ORIGIN

Query Match 0.3%; Score 89.6; DB 25; Length 97;  
Best Local Similarity 94.8%; Pred. No. 0.37;  
Matches 92; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 6698 GTACAAAGTTCTGAGCAGGTCCACCCGACACCGGCATCTCATCCAGGCGCATGGGAT 6757  
|||||  
Db 1 GTACAAAGTTCTGAGCAGGTCCATCCGACACCGGCATCTCTCCAAAGCAATGGGAT 60  
|||||

Qy 6758 CATGAATTCCTTCGTCAACGACATCTTCGAGCGCATC 6794  
|||||  
Db 61 CATGAATTCCTTCGTCAACGACATCTTCGAGCGCATC 97  
|||||

RESULT 11  
AQ028649  
LOCUS

AQ028649 103 bp DNA GSS 30-JUN-1998  
CIT-HSP-2323P12.TR CIT-HSP Homo sapiens genomic clone 2323P12,  
genomic survey sequence.

ACCESSION AQ028649  
VERSION AQ028649.1 GI:3268871  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 103)  
AUTHORS Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Linher, K.,  
Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H.,  
Simon, M. and Venter, J.C.  
Use of a random BAC End Sequence Database for Sequence-Ready Map  
Building (1998)  
Unpublished (1998)

TITLE  
JOURNAL

## COMMENT

Other\_GSSs: CIT-HSP-2323P12.TF  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdams@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
[http://www.tigr.org/tldb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html).  
Seq primer: M13 Reverse  
Class: BAC ends.

## FEATURES

## source

1..103  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2323P12"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2:  
HindIII"

BASE COUNT  
ORIGIN

35 a 27 c 28 g 13 t

## Query Match

Best Local Similarity 0.3%; Score 88.6; DB 94; Length 103;  
Matches 94; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2780 CTAGCACTTTGGAGCCGAGACGCGGATCAGGATCAGGATCGAGACCATCTTG 2839

Db 1 CCAGCACTTAGGAGCCGAGCGCGGCGGATCAGGATCAGGATCGAGACCATCTTG 60

QY 2840 GCTAACACGGTGAACCCCGTTCTACTATAAAATACAAAAAT 2882

Db 61 GCTAACACGGTGAACCCCGCTACTATAAAATACAAAAAT 103

RESULT 12  
AQ076649/c

LOCUS  
DEFINITION  
CIT-HSP-2363C23.TR CIT-HSP Homo sapiens genomic clone 2363C23,  
genomic survey sequence.

ACCESSION  
AQ076649

VERSION  
AQ076649.1 GI:3437833

KEYWORDS  
GSS.

SOURCE  
human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS  
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,  
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and  
Venter,J.C.

TITLE  
Use of a random human BAC End Sequence Database for Sequence-Ready  
Map Building

UNPUBLISHED (1998)

Other\_GSSs: CIT-HSP-2363C23.TF

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:

[http://www.tigr.org/tldb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html).  
Seq primer: M13 Reverse

Class: BAC ends.

## FEATURES

Location/Qualifiers

## source

1..101  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2363C23"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2:  
HindIII"

BASE COUNT  
ORIGIN

12 a 29 c 28 g 32 t

## Query Match

Best Local Similarity 0.3%; Score 88.2; DB 94; Length 101;  
Matches 93; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 7323 GAATGCGTGAACGGGAGCGGAGCTTGCAGTGCAGCGAGATCGCGCATGCGACTCC 7382

Db 101 GAATGCGTGAACCGAGAGACGGAGCTTGCAGTGCAGCGAGATCGCGCAATGCACTCC 42

QY 7383 AGCTGGGTGCAGACGCGAGACTCCGTCACAAAAA 7423

Db 41 AGCTGGGTGCAGACGCGAGACCCCGTCCTCAAAAAA 1

## RESULT 13

AQ264176

LOCUS

DEFINITION  
CITBI-E1-2509A2.TF CITBI-E1 Homo sapiens genomic clone 2509A2,  
genomic survey sequence.

ACCESSION  
AQ264176

VERSION  
AQ264176.1 GI:3792743

KEYWORDS  
GSS.

SOURCE  
human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS  
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,  
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and  
Venter,J.C.

TITLE  
Use of a random human BAC End Sequence Database for Sequence-Ready

Map Building

UNPUBLISHED (1998)

Other\_GSSs: CITBI-E1-2509A2.TF

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:

[http://www.tigr.org/tldb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html).  
Seq primer: M13-21

Class: BAC ends.

Location/Qualifiers

source

1..106

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="2509A2"

/clone\_lib="CITBI-E1"

/sex="male"

/cell\_type="sperm"

/note="Vector: pBelOBAC11; Site\_1: EcoRI; Site\_2: EcoRI;

Caltech Human BAC Library D"

BASE COUNT  
ORIGIN

25 a 30 c 34 g 17 t

## Query Match

0.3%; Score 88.4; DB 105; Length 106;

Best Local Similarity 89.6%; Pred. No. 0.48;  
Matches 95; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2752 CCGGCGGGTGGCTACGCTGTAACTAGCACTTTGGGAGCGGAGCGGGGATC 2811

Db 1 CCGGCGGAGAGTCACGCTGTAACTCCAGCTTTGGGAGCGGAGCGGGTGGATC 60

Qy 2812 ACGAGTCCAGGATCGAGACCATCTTGGCTAACACGCTGAAACCC 2857

Db 61 ACGAGTCCAGGATCGAGACCGCTCTGGCTAACATGGTGAACCC 106

## RESULT 14

AA807640/c

LOCUS

DEFINITION nx08b05.s1 NCI\_CGAP\_GC3 Homo sapiens cDNA clone IMAGE:1255473 3'

similar to contains Alu repetitive element;; mRNA sequence.

ACCESSION AA807640

VERSION AA807640.1 GI:2877108

KEYWORDS EST.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1987)

COMMENT On Jan 19, 1998 this sequence version replaced gi:2151346.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Tissue Procurement: Christoper A. Moskaluk, M.D., Ph.D., Michael

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html

FEATURES

source

1. .103

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1255473"

/clone\_lib="NCI\_CGAP\_GC3"

/tissue\_type="pooled germ cell tumors"

/lab\_host="DH10B"

/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from 3 pooled germ cell tumors, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library is not normalized. Library was constructed by Bento Soares and M. Fatima Bonaudo."

BASE COUNT

ORIGIN

Query Match 0.3%; Score 87.6; DB 38; Length 103;  
Best Local Similarity 91.2%; Pred. No. 0.6;  
Matches 93; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 2766 TCACGCTGTAACTTGGGAGCGGAGCGGATCAGGTCAGGAGA 2825

Db 103 TCACGCTGTAACTTGGGAGCGGAGCGGATCAGGTCAGGAGA 44

Qy 2826 TCAGACCATCTTGGCTAACACGCTGAAACCCGCTTCTACT 2867

Db 43 TCAGACCATCTTGGCTAACACGCTGAAACCCGCTTCTACT 2

## RESULT 15

AW250394

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AW250394

2822460.3prime NIH\_MGC\_7 Homo sapiens cDNA clone IMAGE:2822460 3',

mRNA sequence.

AW250394

AW250394.1 GI:6593387

EST.

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 110)

NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

On May 18, 1998 this sequence version replaced gi:3138342.

Other ESTs: 2822460.5prime

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Tissue Procurement: DCTD/DTF CDNA Library Preparation: Ling

Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E.

Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing

Project Clone distribution: MGC clone distribution information can

be found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html Base Calling / Quality

Scores: PHRED from University of Washington Genome Center

Trimming: cross\_match from University of Washington Genome Center

PHRAP suite. Poly-T Identification: patmatch.pl from Berkeley

Drosophila Genome Project. University of Washington Genome Center:

http://www.genome.washington.edu Low Quality Sequence: 61

Contiguous PHRED high quality bases following vector sequence. Very

Low Quality Sequence: Trace file contained 110 contiguous distinct

peaks following vector sequence. Polyadenylation: Based upon the

presence of a XhoI site followed by a run of 14 or more T residues

polyadenylated.

Plate: LCM9 row: H column: 13

High quality sequence stop: 61.

Location/Qualifiers

1. .110

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2822460"

/clone\_lib="NIH\_MGC\_7"

/tissue\_type="small cell carcinoma"

/cell\_line="MGC3"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: lung; Vector: pOTB7; Site:1: XhoI; Site:2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCAGAG(G). Size-selected &gt;500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT

ORIGIN

Query Match 0.3%; Score 87.6; DB 79; Length 110;  
Best Local Similarity 87.3%; Pred. No. 0.58;  
Matches 96; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 9617 TTTTCTTTTTCAGCGGAGTTTCACACTTTGCCAGGCTGGAGTGAATGGTCGATC 9676

Db 1 TTTTCTTTTTCAGCGGAGTTTCGCTCTATTGTCACAGCTGGAGTGAATGGTCGATC 60

Qy 9677 TCGGCTACCGCAACTCGGCCTCCAGGTTCAAGCAATTCTCCTGCCTC 9726  
||||| ||| | ||||| ||||| || ||||| ||||| ||||| ||||| |||||  
Db 61 TCGGCTCAACTCAACCTCGGCCTCCGGTTTCAAGAGATTCTCCTGCCTC 110

Search completed: June 15, 2000, 19:11:36  
Job time: 110063 sec









CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,862B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-485-862B-65

Query Match 0.3%; Score 80.2; DB 4; Length 105;  
Best Local Similarity 87.1%; Pred. No. 2.5e-09;  
Matches 88; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 TTGTAATTTTATTAGACAGAGGTTTCACTATGTGGCCAGGCTGATCTCAAACTCCTGA 60  
Db 5 TTACATCTTTAGTAGACAGAGGTTTCACTATGTGGCCAGGCTGCTCTCAAACTCCTGA 64

QY 61 CCTCATGATCCCGCTGCGCTTGGCCCTCTCAAAGTGTGGGAT 101  
Db 65 CCTGTGATCCACCAGCCTCGGCCCTCCCAAAAGTGTGGGAT 105

RESULT 5  
US-08-787-739-65  
; Sequence 65, Application US/08787739  
; Patent No. 6027887  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 96  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 369 Pine Street, Suite 610  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/787,739  
FILING DATE: 24-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,049  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/486,756  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/481,658  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,862  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,863  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/487,077  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.4  
TELEPHONE: 415-981-2034  
TELEFAX: 415-981-0332  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-787-739-65

Query Match 0.3%; Score 80.2; DB 5; Length 105;  
Best Local Similarity 87.1%; Pred. No. 2.5e-09;  
Matches 88; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 TTGTAATTTTATTAGACAGAGGTTTCACTATGTGGCCAGGCTGATCTCAAACTCCTGA 60  
Db 5 TTACATCTTTAGTAGACAGAGGTTTCACTATGTGGCCAGGCTGCTCTCAAACTCCTGA 64

QY 61 CCTCATGATCCCGCTGCGCTTGGCCCTCTCAAAGTGTGGGAT 101  
Db 65 CCTGTGATCCACCAGCCTCGGCCCTCCCAAAAGTGTGGGAT 105

RESULT 6  
US-08-481-658B-65/c  
; Sequence 65, Application US/08481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:



MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-486-756A-65

Query Match 0.3%; Score 76.2; DB 4; Length 105;  
Best Local Similarity 82.9%; Pred. No. 2e-08;  
Matches 87; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 2777 ATCCCTAGCACTTTGGGAGCCGCGGATCAGGATCAGGATCGAGACCATC 2836

Db 105 ATCCGAGCACTTTGGGAGCCGCGGATCAGGATCAGGATCGAGACCATC 46

QY 2837 TTGGCTAACACGGTGAACCCCTTCTACTAAATAACAAAAA 2881

Db 45 CTGGCCAATATGTTGAACCCCTGCTCTACTAAAGATGTAATAA 1

## RESULT 9

US-08-485-862B-65/C

; Sequence 65, Application US/08485862B

; Patent No. 5989838

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

; CITY: Tiburon

; STATE: California

; COUNTRY: USA

; ZIP: 94920

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/485,862B

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/477,504

; FILING DATE: 07-JUN-1995

; APPLICATION NUMBER: US 08/260,190

; FILING DATE: 15-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lauder, Leona L.

; REGISTRATION NUMBER: 30,863

; REFERENCE/DOCKET NUMBER: D-0021.3D

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-435-2034

; TELEFAX: 415-435-0727

; INFORMATION FOR SEQ ID NO: 65:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 105 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-485-862B-65

Query Match 0.3%; Score 76.2; DB 4; Length 105;  
Best Local Similarity 82.9%; Pred. No. 2e-08;  
Matches 87; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 2777 ATCCCTAGCACTTTGGGAGCCGCGGATCAGGATCAGGATCGAGACCATC 2836

Db 105 ATCCGAGCACTTTGGGAGCCGCGGATCAGGATCAGGATCGAGACCATC 46

QY 2837 TTGGCTAACACGGTGAACCCCTTCTACTAAATAACAAAAA 2881

Db 45 CTGGCCAATATGTTGAACCCCTGCTCTACTAAAGATGTAATAA 1

## RESULT 10

US-08-787-739-65/c

; Sequence 65, Application US/08787739

; Patent No. 6027887

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 96

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 369 Pine Street, Suite 610

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/787,739

; FILING DATE: 24-JAN-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/485,049

; FILING DATE: 07-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/486,756

; FILING DATE: 07-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/477,504

; FILING DATE: 07-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/481,658

; FILING DATE: 07-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/485,862

; FILING DATE: 07-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/485,863

; FILING DATE: 07-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/487,077

; FILING DATE: 07-JUN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Lauder, Leona L.

; REGISTRATION NUMBER: 30,863

; REFERENCE/DOCKET NUMBER: D-0021.4

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-981-2034

; TELEFAX: 415-981-0332

; INFORMATION FOR SEQ ID NO: 65:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 105 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-787-739-65



```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450.673C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
;
US-08-450-673C-91

Query Match 0.2%; Score 63.8; DB 4; Length 84;
Best Local Similarity 85.5%; Pred. NO. 1.1e-05;
Matches 71; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 2767 CACGCCTGTAATCCTAGCACACCTTTGGAGCGCGGATCACAGGTCAGAGAT 2826
Db 83 CACGCTTGTATCCAGCACACCTTTGGAGGCTGAGCGCGGATCACAGGTCAGGAGTT 24

Qy 2827 CGAGACCATCTTGGCTAACACGG 2849
Db 23 CGACACCAGCCTGATGAACATGG 1

RESULT 14
PCT-US95-17111A-91/c
; Sequence 91, Application PC/TUS9517111A
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and
; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 69:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 76 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
;
US-08-454-557C-69

Query Match 0.2%; Score 62.8; DB 3; Length 76;
Best Local Similarity 97.0%; Pred. NO. 1.8e-05;
Matches 64; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 8680 ACCATGCCCGCTAATTTTGTATTTTAGTAGACACAGGGTTTACCCTGTGGCCAGG 8739
Db 1 ACCACGCCCGCTAATTTTGTATTTTAGTAGACACAGGGTTTACCCTGTGGCCAGG 60

Qy 8740 ATGCTC 8745
Db 61 ATGCTC 66
```

```
;
; TOPOLOGY: both
;
PCT-US95-17111A-91

Query Match 0.2%; Score 63.8; DB 6; Length 84;
Best Local Similarity 85.5%; Pred. NO. 1.1e-05;
Matches 71; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 2767 CACGCCTGTAATCCTAGCACACCTTTGGAGCGCGGATCACAGGTCAGAGAT 2826
Db 83 CACGCTTGTATCCAGCACACCTTTGGAGGCTGAGCGCGGATCACAGGTCAGGAGTT 24

Qy 2827 CGAGACCATCTTGGCTAACACGG 2849
Db 23 CGACACCAGCCTGATGAACATGG 1

RESULT 15
US-08-454-557C-69
; Sequence 69, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454.557C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 69:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 76 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
;
US-08-454-557C-69

Query Match 0.2%; Score 62.8; DB 3; Length 76;
Best Local Similarity 97.0%; Pred. NO. 1.8e-05;
Matches 64; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 8680 ACCATGCCCGCTAATTTTGTATTTTAGTAGACACAGGGTTTACCCTGTGGCCAGG 8739
Db 1 ACCACGCCCGCTAATTTTGTATTTTAGTAGACACAGGGTTTACCCTGTGGCCAGG 60

Qy 8740 ATGCTC 8745
Db 61 ATGCTC 66
```

Search completed: June 16, 2000, 00:17:42  
Job time: 127392 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2000, 00:11:20 ; Search time 17971 Seconds  
(without alignments)  
-1569.858 Million cell updates/sec

Title: US-08-852-495C-1\_COPY\_112000\_141000  
Perfect score: 29001  
Sequence: 1 TTTTTCACACTCTTCTTCAG.....AGAGTGTTCACCTCTAGGA 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 370290

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pl1.\*
- 8: gb\_pl2.\*
- 9: gb\_pr1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_sts.\*
- 14: gb\_sy.\*
- 15: gb\_un.\*
- 16: gb\_vl.\*
- 17: em\_fun.\*
- 18: em\_hum1.\*
- 19: em\_hum2.\*
- 20: em\_in.\*
- 21: em\_om.\*
- 22: em\_or.\*
- 23: em\_ov.\*
- 24: em\_pat.\*
- 25: em\_ph.\*
- 26: em\_pl.\*
- 27: em\_ro.\*
- 28: em\_sts.\*
- 29: em\_sy.\*
- 30: em\_un.\*
- 31: em\_vl.\*
- 32: gb\_htg1.\*
- 33: gb\_htg2.\*
- 34: gb\_in1.\*
- 35: gb\_in2.\*
- 36: em\_ba1.\*
- 37: em\_ba2.\*
- 38: em\_hum3.\*
- 39: em\_hum4.\*
- 40: gb\_pr4.\*
- 41: gb\_htg3.\*
- 42: gb\_htg4.\*
- 43: gb\_htg5.\*
- 44: gb\_htg6.\*

- 45: gb\_htg7.\*
- 46: em\_htg1.\*
- 47: em\_htg2.\*
- 48: em\_htg3.\*
- 49: em\_hum5.\*
- 50: gb\_pl3.\*
- 51: gb\_pr5.\*
- 52: gb\_htg8.\*
- 53: gb\_htg9.\*
- 54: gb\_htg10.\*
- 55: gb\_htg11.\*
- 56: gb\_htg12.\*
- 57: gb\_htg13.\*
- 58: gb\_htg14.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	99	0.3	108	10	HSLDLRN2
2	90.2	0.3	107	9	HUMALCE162
3	89.4	0.3	108	10	HSLDLRD1
4	89.4	0.3	108	10	HSLDLRD2
5	86.8	0.3	108	10	HSLDLRN2
6	85.4	0.3	103	9	HUMALCE221
7	83.2	0.3	108	10	HSLDLI12
8	82.6	0.3	108	11	HSU67803
9	81.2	0.3	107	9	HUMALCE162
10	79.8	0.3	103	9	HUMALCE221
11	79.4	0.3	108	10	HSLDLRD1
12	79.4	0.3	108	10	HSLDLRD2
13	76	0.3	103	13	HS8IC8R
14	75.6	0.3	110	11	HSU67807
15	75.2	0.3	103	13	HS8IC8R
16	74.4	0.3	104	9	HUMALCE272
17	74.2	0.3	108	9	HUMD1D03M5
18	74.4	0.3	108	11	HSU67808
19	73.6	0.3	97	9	HUMDLRA2
20	73.4	0.3	110	11	HSU67807
21	71.4	0.2	108	11	HSU67804
22	70.8	0.2	99	13	HUMUT7692A
23	70.8	0.2	108	10	HSLDLI12
24	70	0.2	95	13	HUMUT8002B
25	70.2	0.2	108	13	G32614
26	69.8	0.2	101	10	S79560
27	69.6	0.2	107	11	HSU67806
28	69.4	0.2	108	13	G43535
29	69	0.2	90	9	HUMDLRFL
30	69.2	0.2	91	13	HUMUT8164A
31	68.4	0.2	79	10	S73203
32	68.2	0.2	101	10	S79560
33	67.8	0.2	95	10	HSSTHPK1B
34	68	0.2	108	13	G43535
35	67.2	0.2	80	9	HUMBRKFAE
36	66.6	0.2	91	13	HUMUT8164A
37	66.8	0.2	94	9	HUMHGAL
38	66.2	0.2	100	9	HUMGALNSA
39	66.2	0.2	100	9	HUMGALNS
40	65.6	0.2	97	9	HUMDLRA2
41	65.2	0.2	108	9	HUMD1D03M5
42	64.4	0.2	90	10	HSU19407
43	64.6	0.2	100	10	HSLAS27
44	64.4	0.2	110	9	HUMALCE43
45	64	0.2	90	9	HUMDLRFL

ALIGNMENTS

```

RESULT 1
HSLDLR2  HSLDLR2  108 bp  DNA  PRI  20-MAY-1992
LOCUS    Human LDL-receptor gene intron 14 fragment (normal gene).
DEFINITION
ACCESSION X05250
VERSION   X05250.1 GI:34337
KEYWORDS  Alu repetitive sequence; low density lipoprotein receptor.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
           Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS   Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
           Williamson,R., and Humphries,S.
TITLE     Unequal crossing-over between two alu-repetitive DNA sequences in
           the low-density-lipoprotein-receptor gene. A possible mechanism for
           the defect in a patient with familial hypercholesterolaemia
JOURNAL   Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE   87161901
COMMENT   See X05252 for deletion junction
           Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES
   source
       location/Qualifiers
           1..108
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           1..108
           /note="intron XIV fragment"
BASE COUNT 28 a 23 c 39 g 18 t
ORIGIN
intron
Query Match 0.3%; Score 99; DB 10; Length 108;
Best Local Similarity 95.3%; Pred. No. 1e-07;
Matches 102; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 25347 ACAAAATCAGCAGCGCTGGTGGCATGTGCTGTAATCCAGCTACTCAGGAGCTGAG 25406
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 ACAAAATAGCAGCGCTGGTGGCATGTGCTGTAATCCAGCTACTCAGGAGCTGAG 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Qy 25407 GCAAGAGATTCCTGAACCCAGGAGCGGAGGTTCAGTGCAGTGCAGCGCA 25453
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 61 GCAGGAGAATTCCTGAACCCAGGAGCGGAGGTTCAGTGCAGTGCAGCGCA 107
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 2
HUMALCE162/c HUMALCE162 107 bp ss-RNA PRI 15-APR-1994
LOCUS    Human carcinoma cell-derived Alu RNA transcript, clone CE162.
DEFINITION
ACCESSION M87924
VERSION   M87924.1 GI:174871
KEYWORDS  Alu repeat.
SOURCE    Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 107)
AUTHORS   Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE     Alu RNA transcripts in human embryonal carcinoma cells. Model of
           post-transcriptional selection of master sequences
JOURNAL   J. Mol. Biol. (1992) in press
FEATURES
   source
       location/Qualifiers
           1..107
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /cell_line="NTER2D1"
           /dev_stage="embryo"
           /sex="male"
           /tissue_type="carcinoma"
BASE COUNT 28 a 30 c 35 g 14 t
ORIGIN

```

```

Query Match 0.3%; Score 90.2; DB 9; Length 107;
Best Local Similarity 92.2%; Pred. No. 3.4e-06;
Matches 95; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 20076 TTTTGTGAGACGGAGTCTGCTGTGTCGCCAGGCTGGAATGCAGTGCACAAATCTCGC 20135
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 107 TTTTGTGAGACGGAGTCTGCTGTGTCGCCAGGCTGGAATGCAGTGCAGTGCAGTCTCGC 48
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Qy 20136 TCACTGCACACTCCGCTCCCGGATTACGCCCATCTCCTGCC 20178
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 47 TCACTGCACAGTCCGCTCCCGGTTTACGCCCATCTCTCTGCC 5

RESULT 3
HSLDLR1/c HSLDLR1 108 bp DNA PRI 20-MAY-1992
LOCUS    Human LDL-receptor mutated gene with intron 12 deletion junction.
DEFINITION
ACCESSION X05249
VERSION   X05249.1 GI:34335
KEYWORDS  Alu repetitive sequence; low density lipoprotein receptor.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
           Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS   Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
           Williamson,R., and Humphries,S.
TITLE     Unequal crossing-over between two alu-repetitive DNA sequences in
           the low-density-lipoprotein-receptor gene. A possible mechanism for
           the defect in a patient with familial hypercholesterolaemia
JOURNAL   Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE   87161901
COMMENT   *source: hypercholesterol aemia
           See X05248 for corresponding normal gene sequence
           In the defective LDL-receptor gene the deletion occurred between two
           alu-repetitive sequences, that are in the same direction, the
           deletion eliminates exons 13 and 14 and changes the reading frame
           of the resulting spliced mRNA.
           Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.
FEATURES
   source
       location/Qualifiers
           1..108
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /cell_type="blood leukocytes from a patient with familial"
   misc_feature
       1..108
       /note="deletion junction region intron 12/ intron 15"
BASE COUNT 20 a 40 c 20 g 28 t
ORIGIN

Query Match 0.3%; Score 89.4; DB 10; Length 108;
Best Local Similarity 89.7%; Pred. No. 4.6e-06;
Matches 96; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 25347 ACAAAATCAGCAGCGCTGGTGGCATGTGCTGTAATCCAGCTACTCAGGAGCTGAG 25406
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 108 ACAAAATAGCAGCGCTGGTGGCATGTGCTGTAATCCAGCTACTCAGGAGCTGAG 49
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Qy 25407 GCAAGAGATTCCTGAACCCAGGAGCGGAGGTTCAGTGCAGTGCAGCGCA 25453
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 48 GCAGGAAAATGGTTGAACCCAGGAGCGGAGGTTCAGTGCAGCGCA 2

RESULT 4
HSLDLR2  HSLDLR2  108 bp  DNA  PRI  20-MAY-1992
LOCUS    Human LDL-receptor mutated gene with intron 14 deletion junction.
DEFINITION
ACCESSION X05251
VERSION   X05251.1 GI:34336
KEYWORDS  Alu repetitive sequence; low density lipoprotein receptor.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

```

JOURNAL  
MEDLINE 87161901

COMMENT see x05249 for deletion junction  
Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.

```
FEATURES
  source      1..108
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              complement(<1..65)
              /note="Alu repeat"
  intron      1..108
              /note="intron XII fragment"
  BASE COUNT 21 a 38 c 20 g 29 t
  ORIGIN
      0.3%; Score 83.2; DB 10; Length 108;
      Best Local Similarity 87.5%; Pred. No. 5.4e-05;
      Matches 91; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 25350 AAAATCAGCCAGCGTGGTGGCATGTGCTGTAATCCAGCTACTCAGGAGCTGAGGCA 25409
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105 AAAATCAGCCGGCGTGGTGGCACATCTTGTATCCAGCTACTAAGGAGCTGAGGCA 46
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 25410 AGAGAATTGCTTGAACCCAGGAGGCGGAGGTTGCGAGTGAGCGGA 25453
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 45 GGAATAATGGTTTGAACCCAGGAGGAGAGGTTGTGGTGAGGCGA 2
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
HSU67803/c 108 bp RNA PRI 01-AUG-1997
LOCUS HSU67803 Human small cytoplasmic Alu transcript.
DEFINITION U67803
ACCESSION U67803
VERSION U67803.1 GI:2289917
KEYWORDS Alu.
SOURCE human.
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
    Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1 (bases 1 to 108)
  AUTHORS Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.
  TITLE CDNAs derived from primary and small cytoplasmic Alu (scAlu)
    transcripts
  JOURNAL J. Mol. Biol. 271 (2), 222-234 (1997)
  MEDLINE 97415756
  REFERENCE 2 (bases 1 to 108)
  AUTHORS Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.
  TITLE Direct Submission
  JOURNAL Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The
    Children's Hospital of Philadelphia, 1004F Abramson Research
    Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA

FEATURES
  source      1..108
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="TscAlu2"
  repeat_region 1..108
              /note="scAlu"
              /rpt_family="Alu"
              /rpt_type="dispersed"
  BASE COUNT 23 a 39 c 30 g 16 t
  ORIGIN
      0.3%; Score 82.6; DB 11; Length 108;
      Best Local Similarity 90.7%; Pred. No. 6.9e-05;
      Matches 88; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 20246 GTAGACAGGGTTTACCGTGTACCGGGGATGTCGATCTCCTGACCTCATGATCT 20305
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 97 GTAGACAGGGGTTTACACCTTCTTAGCCAGGATGTCGATCTCCTGACCTCGTATCC 38
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 20306 GCCCACCCTCAGCTCCCAAGTGCTAGGATCACAGC 20342
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 37 GCCCGCCTCGGCCTCCCAAGTGCTGGGATTACAGC 1
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
HUMALCE162 107 bp ss-RNA PRI 15-APR-1994
LOCUS HUMALCE162 Human carcinoma cell-derived Alu RNA transcript, clone CE162.
DEFINITION M87924
ACCESSION M87924
VERSION M87924.1 GI:174871
KEYWORDS Alu repeat.
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
    Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1 (bases 1 to 107)
  AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
  TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
    post-transcriptional selection of master sequences
  JOURNAL J. Mol. Biol. (1992) In press
  FEATURES
    Location/Qualifiers
      1..107
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /cell_line="NTERa2D1"
        /dev_stage="embryo"
        /sex="male"
        /tissue_type="carcinoma"
  BASE COUNT 28 a 30 c 35 g 14 t
  ORIGIN
      0.3%; Score 81.2; DB 9; Length 107;
      Best Local Similarity 87.3%; Pred. No. 0.00012;
      Matches 89; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 25406 GCGAAGAGAATTGTTGAACCCAGGAGCGGAGGTTGCGAGTGAGCCGAAATCGCGCCACT 25465
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5 GCGAAGAGAATTGCGTGAACCCGAGCGGAGGAGTGTGCGAGTGAGCGGAGATCGCGCCACT 64
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 25466 GCATCCAGCTGGTGAACAGAGCAAGGCTCTGTTTCAAAAA 25507
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 GCATCCAGCTGGCGAGAGCGAGACTCCGTCCTCAAAA 106
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
HUMALCE221/c 103 bp ss-RNA PRI 15-APR-1994
LOCUS HUMALCE221 Human carcinoma cell-derived Alu RNA transcript, clone CE221.
DEFINITION M87896
ACCESSION M87896
VERSION M87896.1 GI:174874
KEYWORDS Alu repeat.
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
    Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1 (bases 1 to 103)
  AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
  TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
    post-transcriptional selection of master sequences
  JOURNAL J. Mol. Biol. (1992) In press
  FEATURES
    Location/Qualifiers
      1..103
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /cell_line="NTERa2D1"
        /dev_stage="embryo"
        /sex="male"
        /tissue_type="carcinoma"
  BASE COUNT 25 a 27 c 33 g 18 t
  ORIGIN
      0.3%; Score 79.8; DB 9; Length 103;
      Query Match
```

```
Best Local Similarity 87.9%; Pred. No. 0.00021;
Matches 87; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 20110 CTGGAATGCAGTGGCAACAATCTGGCTCACTGCAACCTCGCCCTCCGGGATTACGGCAT 20169
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 103 CTGGAGTGAATGCAGGATCTGGCTCACTGCAACCTCGCCCTCCGGGTTCAAGCGAT 44
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 20170 TCTCCTCGCTCAACCTCCCGAGTAGCTGGGACACAGGC 20208
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 43 TCTCCTCGCTTAGTCTCCCGTAGCTGGGATTACAGGC 5
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
LOCUS HSLDLRD1 108 bp DNA PRI 20-MAY-1992
DEFINITION Human LDL-receptor mutated gene with intron 12 deletion junction.
ACCESSION X05249
VERSION X05249.1 GI:34335
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 108)
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="blood leukocytes from a patient with familial"
/notes="intron XIV fragment"
28 a 20 c 40 g 20 t

Intron 1..108
BASE COUNT 28 a 20 c 40 g 20 t
ORIGIN

Query Match 0.3%; Score 79.4; DB 10; Length 108;
Best Local Similarity 84.8%; Pred. No. 0.00025;
Matches 89; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 20131 TCGGCTCACTGCAACCTCCGGCTCCCGGATTACGGCATTTCTCTGCTCAACCTCCCGA 20190
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 107 TCGCCTCACCACACACTCTGCCTCTGGGTTCAACCATTTTCTGCTCAGCCTCCCGA 48
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 20191 GTAGCTGGGACCAACAGCGGCCGCCACGACCCAGCTAATTTT 20235
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 47 GTAGCTGGGATTACAGCAGCTGCACGAGCGCTGGCTAATTTT 3
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

FEATURES
source 1..108
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="blood leukocytes from a patient with familial"
Location/Qualifiers
1..108
/notes="intron XIV fragment"
28 a 20 c 40 g 20 t

Intron 1..108
BASE COUNT 28 a 20 c 40 g 20 t
ORIGIN

Query Match 0.3%; Score 79.4; DB 10; Length 108;
Best Local Similarity 84.8%; Pred. No. 0.00025;
Matches 89; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 20131 TCGGCTCACTGCAACCTCCGGCTCCCGGATTACGGCATTTCTCTGCTCAACCTCCCGA 20190
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 107 TCGCCTCACCACACACTCTGCCTCTGGGTTCAACCATTTTCTGCTCAGCCTCCCGA 48
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 20191 GTAGCTGGGACCAACAGCGGCCGCCACGACCCAGCTAATTTT 20235
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 47 GTAGCTGGGATTACAGCAGCTGCACGAGCGCTGGCTAATTTT 3
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
HS8IC8R/c 103 bp DNA STS 05-SEP-1991
LOCUS Human sequence tagged site 8IC8R DNA from 19q13.
DEFINITION X57789
ACCESSION X57789
VERSION X57789.1 GI:23938
KEYWORDS STS; myotonic dystrophy.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 103)
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19q13"
/germline
/clone_lib="YAC library: ICI"
/clone="8IC8"
29 a 28 c 23 g 22 t 1 others

REFERENCE
AUTHORS Butler,R., Riley,J.H., Ogilvie,D.J., Anand,R., Buxton,J.,
TITLE Davies,J., Johnson,K. and Markham,A.F.
JOURNAL Submitted (12-FEB-1991) F.L. Aldridge, ICI Pharmaceuticals,
MEDLINE Alderley Park, Macclesfield, Cheshire, SK10 4TG, UK
COMMENT 2 (bases 1 to 103)
FEATURES
source 1..103
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19q13"
/germline
/clone_lib="YAC library: ICI"
/clone="8IC8"
29 a 28 c 23 g 22 t 1 others

Best Local Similarity 87.9%; Pred. No. 0.00021;
Matches 87; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 20110 CTGGAATGCAGTGGCAACAATCTGGCTCACTGCAACCTCGCCCTCCGGGATTACGGCAT 20169
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 103 CTGGAGTGAATGCAGGATCTGGCTCACTGCAACCTCGCCCTCCGGGTTCAAGCGAT 44
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 20170 TCTCCTCGCTCAACCTCCCGAGTAGCTGGGACACAGGC 20208
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 43 TCTCCTCGCTTAGTCTCCCGTAGCTGGGATTACAGGC 5
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
LOCUS HSLDLRD1 108 bp DNA PRI 20-MAY-1992
DEFINITION Human LDL-receptor mutated gene with intron 12 deletion junction.
ACCESSION X05249
VERSION X05249.1 GI:34335
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 108)
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="blood leukocytes from a patient with familial"
/notes="deletion junction region intron 12/ intron 15"
20 a 40 c 20 g 28 t

misc_feature 1..108
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="blood leukocytes from a patient with familial"
Location/Qualifiers
1..108
/notes="deletion junction region intron 12/ intron 15"
20 a 40 c 20 g 28 t

BASE COUNT 20 a 40 c 20 g 28 t
ORIGIN

Query Match 0.3%; Score 79.4; DB 10; Length 108;
Best Local Similarity 84.8%; Pred. No. 0.00025;
Matches 89; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 20131 TCGGCTCACTGCAACCTCCGGCTCCCGGATTACGGCATTTCTCTGCTCAACCTCCCGA 20190
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 TCGCCTCACCACACACTCTGCCTCTGGGTTCAACCATTTTCTGCTCAGCCTCCCGA 61
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 20191 GTAGCTGGGACCAACAGCGGCCGCCACGACCCAGCTAATTTT 20235
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 GTAGCTGGGATTACAGCAGCTGCACGAGCGCTGGCTAATTTT 106
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
HSLDLRD2/c 108 bp DNA PRI 20-MAY-1992
LOCUS Human LDL-receptor mutated gene with intron 14 deletion junction.
DEFINITION X05251
ACCESSION X05251
VERSION X05251.1 GI:34336
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
```



\_\_\_\_\_







```
RESULT 2
X12095 ID X12095 standard; DNA; 108 BP.
AC X12095;
DT 30-MAR-1999 (first entry)
DE Human biallelic polymorphic DNA fragment TIGR-A003M18a.
KW polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medication;
KW treatment; marker; ss.
OS Homo sapiens.
PN WO9820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; U20313.
PR 06-NOV-1996; US-030455.
PA (WHEED ) WHITEHEAD INST. BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
DR WPI; 98-286974/25.
PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
PS Claim 1: Page 219; 310pp; English.
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC X09121-X10268. The base occupying the polymorphic site is indicated by
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC segments can also be used to produce medicaments for the treatment or
CC prophylaxis of such diseases.
CC Sequence 108 BP; 19 A; 23 C; 28 G; 37 T;
SQ
```

```
Query Match 0.2%; Score 68.6; DB 1; Length 108;
Best Local Similarity 82.8%; Pred. No. 0.015;
Matches 77; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

QY 10310 TGTATTTTGTAGACGGGTTTTCACATGTTGCCAGGCTGCTCTCAAACTCCTGCAC 10369
Db 1 TGTCTTTTGTAGAGATGAGTGTCTTCTCTGTGTCAGGATGCTCTCGAACTCCTGCAC 60

QY 10370 CTCAGGTGATCACCCTCGCTCGGCTCCCAAAA 10402
Db 1 TTTCAAGTGATCGCTGCTCGCTCGGCTCCCAAAA 93

RESULT 3
T24892 ID T24892 standard; cDNA to mRNA; 100 BP.
AC T24892;
DT 05-NOV-1996 (first entry)
DE Human gene signature HUMGS06998.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATSU/) MATSUBARA K.
PA (OKUBA/) OKUBO K.
```

```
Query Match 0.2%; Score 63.2; DB 1; Length 100;
Best Local Similarity 77.9%; Pred. No. 0.1;
Matches 74; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 28738 ATCACTTGAACCTGGAGGCGAGAGTTGTCAGTGAGGGAGATGGCCACTGCACCTCCAG 28797
Db 2 ATCGCTTGAACCTGGAGGCGAGAGTTTGCATNAGCTGAGATTGCACCTTGCACCTCCG 61

QY 28798 CCTGAGCAACACAGCGAGACTCTGTCACAAAAA 28832
Db 62 CCTGGGTGACAGAGTGACACTCTGTTGAAACAAA 96

RESULT 4
T25009 ID T25009 standard; cDNA to mRNA; 108 BP.
AC T25009;
DT 07-NOV-1996 (first entry)
DE Human gene signature HUMGS07131.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATSU/) MATSUBARA K.
PA (OKUBA/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1748; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
CC Sequence 100 BP; 28 A; 22 C; 25 G; 22 T;
SQ
```

```
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1720; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
CC Sequence 100 BP; 28 A; 22 C; 25 G; 22 T;
SQ
```

```
Query Match 0.2%; Score 63.2; DB 1; Length 100;
Best Local Similarity 77.9%; Pred. No. 0.1;
Matches 74; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 28738 ATCACTTGAACCTGGAGGCGAGAGTTGTCAGTGAGGGAGATGGCCACTGCACCTCCAG 28797
Db 2 ATCGCTTGAACCTGGAGGCGAGAGTTTGCATNAGCTGAGATTGCACCTTGCACCTCCG 61

QY 28798 CCTGAGCAACACAGCGAGACTCTGTCACAAAAA 28832
Db 62 CCTGGGTGACAGAGTGACACTCTGTTGAAACAAA 96

RESULT 4
T25009 ID T25009 standard; cDNA to mRNA; 108 BP.
AC T25009;
DT 07-NOV-1996 (first entry)
DE Human gene signature HUMGS07131.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATSU/) MATSUBARA K.
PA (OKUBA/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1748; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
```



PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues

PS Claim 1; Page 1720; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in T19001-T26837 and which is able to hybridise to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost

CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library

CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be

CC determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.

CC Sequence 100 BP; 28 A; 22 C; 25 G; 22 T;

SQ

	Query Match	0.28;	Score 59.2;	DB 1;	Length 100;
	Best Local Similarity	73.7;	Pred. No. 0.41;	26;	Indels 0;
	Matches 73;	Conservative	0;	Mismatches	0;
Qy	20071	TTTCTTTTCGAGACGGAGCTTGTCTCTGCGCCCGAGCTGGAAATCAGTCGGCAACAATC	20130		
Db	100	TTTTTTTTTTTCAACACAGAGTGTCACTGTGCACCCAGGCGNGAGTGCAANGGTGCAATC	41		
Qy	20131	TCGGCTCACTGCAACCTCCGCTCCCGGATTCAGGCCAT	20169		
Db	40	TCAGCTNATTCGAATTCGTGCTCCCGAGTTCAAGCGAT	2		

RESULT	8	
T25009/c		
ID	T25009	standard; cDNA to mRNA; 108 BP.
AC	T25009;	
DT	07-NOV-1996	(first entry)
DE	Human gene signature HUNG507131.	
KE	Gene signature; messenger RNA; mRNA; relative abundance; frequency;	
KW	human; cloning; mapping; non-biased library; diagnosis; detection;	
OS	cell typing; abnormal cell function; ss.	
KS	Homo sapiens.	
PN	W09514772-A1.	
PD	01-JUN-1995.	
PF	11-NOV-1994; J01916.	
PR	12-NOV-1993; JP-355504.	
PA	(MATS/) MATSUBARA K.	
PA	(OKUB/) OKUBO K.	
PI	Matsubara K, Okubo K;	
DR	WPI; 95-206931/27.	
PT	Identifying gene signatures in 3'-directed human cDNA library - e.g.	
PT	for diagnosis of abnormal cell function, by preparing cDNA that	
PT	reflects relative abundance of corresp. mRNA in specific human	
PT	tissues	

PS Claim 1; Page 1748; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in T19001-T26837 and which is able to hybridize to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost

CC all the 3'-oriented cDNAs hybridize with specific mRNAs. Each library

CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be

CC determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.

Seq	Sequence	108 BP;	34 A;	31 C;	26 G;	15 T;
	Query Match		0.28;	Score 58.6;	DB 1;	Length 108;
	Best Local Similarity		71.0%;	Pred. No. 0.51;		
	Matches	76;	Conservative	0;	Mismatches	31;
					Indels	0;
					Gaps	0;
Qy	20063	TTTTTTTTTTCTTTTCTGAGCAGGAGCTTGCTCTGCTCGCCCGAGGCTGGAATGCAGTG	20122			
Db	108	TTTGTTGTTGTTGTTGTTTCAACAGGCTCTGCTCTCACTCAGGCTGGAATNCAGTG	49			
Qy	20123	GCACAACTCGGCTCACTGCACACCTCCGCCCTCCCGGATTCAGGCCAT	20169			
Db	48	CGGTGACCATGGCTCACTGCAGCCTTGCCCTCAATGGGCTCAGGGCAT	2			

RESULT	9	
T21566		
ID	T21566 standard; cDNA to mRNA; 87 BP.	
AC	T21566;	
DT	03-AUG-1996 (first entry)	
DE	Human gene signature HUMGS02944.	
KW	Gene signature; messenger RNA; mRNA; relative abundance; frequency;	
KW	human; cloning; mapping; non-biased library; diagnosis; detection;	
KW	cell typing; abnormal cell function; ss.	
OS	Homo sapiens.	
PN	W09514772-A1.	
PD	01-JUN-1995.	
PF	11-NOV-1994; J01916.	
PR	12-NOV-1993; JP-355504.	
PA	(MATS/) MATSUBARA K.	
PA	(OKUB/) OKUBO K.	
PI	Matsubara K, Okubo K;	
DR	WPI: 95-206931/27.	
PT	Identifying gene signatures in 3'-directed human cDNA library - e.g.	
PT	for diagnosis of abnormal cell function, by preparing cDNA that	
PT	reflects relative abundance of corresp. mRNA in specific human	
PT	tissues	
PS	Claim 1: Page 914; 2245pp; Japanese.	
CC	A single-stranded DNA (or its complementary strand or the corresp.	
CC	double-stranded DNA) which comprises one of the 7837 "GS" sequences	
CC	given in T19001-T26837 and which is able to hybridise to part of	
CC	human genomic DNA, cDNA or mRNA is claimed. The GS (gene signature)	
CC	sequences were obtained from 3'-directed cDNA libraries prepared	
CC	from various human tissues; synthesis of cDNA was initiated from the	
CC	3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-	
CC	untranslated sequence is unique to a particular mRNA species, almost	
CC	all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library	
CC	is constructed so as to reflect accurately the relative abundance of	
CC	different mRNAs in the particular tissue from which it was derived.	
CC	The appearance frequency of a given GS in a cDNA library can be	
CC	determined (esp. using primers and probes derived from the GS	
CC	sequences) as a means of diagnosing abnormal cell function or for	
CC	recognising different cell types.	
SQ	Sequence 87 BP; 35 A; 21 C; 16 G; 13 T;	

	Query Match	0.2%	Score 58;	DB 1;	Length 87;
	Best Local Similarity	79.8%	Pred. No. 0.63;		
	Matches 67;	Conservative 0;	Mismatches 17;	Indels 0;	Gaps 0;
Qy	20924	GATCAGTTGAGTCACAGAGTTTGAGACACAGCGCTGGTCAACATGGGGAACCTCATCTCTA	20983		
Db	1	GATCGCTTGACCCAGGAGTTTAAACACGCCGCGAGGAACATGGCGAAACCCCATCTTTA	60		
Qy	20984	CAAAAATAAAAAATTTGTCAGG	21007		
Db	61	CAAAAAATACAGAAATNAGCCAAAG	84		

RESULT 10  
T21566/c  
ID T21566 standard; cDNA to mRNA; 87 BP.



```
Query Match      0.2%; Score 57.4; DB 1; Length 97;
Best Local Similarity 75.3%; Pred. No. 0.78;
Matches 70; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 7593 GATCTGGCCACTCGGCTCCGAGAGTGTGGGATTACAGGTGAGGCGCCGAGA 7652
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 GATCTGCCACCTNGCGCTCCAGAGTGTGGGATTACAGGCGATGAGCCGCGG 60

QY 7653 CCTGGACTTGTCTCTGTTTCATCAGTCCTTC 7685
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 NCTGTACTAAGTCTTTTCTTTTAAATTCCTC 93

RESULT 13
ID X12087/c
AC X12087 standard; DNA; 100 BP.
DE 30-MAR-1999 (first entry)
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW treatment; marker; ss.
OS Homo sapiens.
PN W09820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; U20313.
PR 06-NOV-1996; US-030455.
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
DR WPI; 98-286974/25.
PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
PS Claim 1; Page 218; 310pp; English.
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC X09121-X10268. The base occupying the polymorphic site is indicated by
CC the appropriate IUPAC-TUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC segments can also be used to produce medicaments for the treatment or
CC prophylaxis of such diseases.
SQ Sequence 100 BP; 21 A; 25 C; 22 G; 31 T;

Query Match      0.2%; Score 57; DB 1; Length 100;
Best Local Similarity 72.7%; Pred. No. 0.9;
Matches 72; Conservative 1; Mismatches 26; Indels 0; Gaps 0;

QY 25232 GTGGCTCACCTGTAATCCAGCACTTTGGAGGCCAAGGTAAAGCATCACTTGAGT 25291
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99 GTGACTCACCTATAATCTCTGGCACCTTTGGAGGCTTAGGAAGGAGGATTGTTTGAAC 40

QY 25292 CAGGAGTTAGACACAGCTCTGCCCAACATAGTGAACATC 25330
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 39 CAGGAGCTCAAGACCATCTCTGGAAACATAGCAAGATC 1

RESULT 14
ID X12085/c
AC X12085 standard; DNA; 100 BP.
DE 30-MAR-1999 (first entry)
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW treatment; marker; ss.
OS Homo sapiens.
PN W09820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; U20313.
PR 06-NOV-1996; US-030455.
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
DR WPI; 98-286974/25.
PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
PS Claim 1; Page 218; 310pp; English.
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC X09121-X10268. The base occupying the polymorphic site is indicated by
CC the appropriate IUPAC-TUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC segments can also be used to produce medicaments for the treatment or
CC prophylaxis of such diseases.
SQ Sequence 100 BP; 21 A; 25 C; 22 G; 31 T;

Query Match      0.2%; Score 57; DB 1; Length 100;
Best Local Similarity 72.7%; Pred. No. 0.9;
Matches 72; Conservative 1; Mismatches 26; Indels 0; Gaps 0;

QY 25232 GTGGCTCACCTGTAATCCAGCACTTTGGAGGCCAAGGTAAAGCATCACTTGAGT 25291
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99 GTGACTCACCTATAATCTCTGGCACCTTTGGAGGCTTAGGAAGGAGGATTGTTTGAAC 40

QY 25292 CAGGAGTTAGACACAGCTCTGCCCAACATAGTGAACATC 25330
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 39 CAGGAGCTCAAGACCATCTCTGGAAACATAGCAAGATC 1

RESULT 14
ID X12085/c
AC X12085 standard; DNA; 100 BP.
DE 30-MAR-1999 (first entry)
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW treatment; marker; ss.
OS Homo sapiens.
PN W09820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; U20313.
PR 06-NOV-1996; US-030455.
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
DR WPI; 98-286974/25.
PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
PS Claim 1; Page 218; 310pp; English.
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC X09121-X10268. The base occupying the polymorphic site is indicated by
CC the appropriate IUPAC-TUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC segments can also be used to produce medicaments for the treatment or
CC prophylaxis of such diseases.
SQ Sequence 100 BP; 21 A; 25 C; 22 G; 31 T;
```

```
AC X12085;
DE 30-MAR-1999 (first entry)
KW Human biallelic polymorphic DNA fragment EST98276c.
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW treatment; marker; ss.
OS Homo sapiens.
PN W09820165-A2.
PD 14-MAY-1998.
PF 05-NOV-1997; U20313.
PR 06-NOV-1996; US-030455.
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI Hudson T, Lander ES, Wang D;
DR WPI; 98-286974/25.
PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
PS Claim 1; Page 218; 310pp; English.
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic
CC markers which have been isolated using the primers represented in
CC X09121-X10268. The base occupying the polymorphic site is indicated by
CC the appropriate IUPAC-TUB ambiguity code. These fragments can be used in
CC methods for determining polymorphic forms in an individual for use in
CC e.g. forensics, paternity testing or for phenotypic typing for diseases
CC such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial
CC hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC segments can also be used to produce medicaments for the treatment or
CC prophylaxis of such diseases.
SQ Sequence 100 BP; 22 A; 25 C; 22 G; 30 T;

Query Match      0.2%; Score 57; DB 1; Length 100;
Best Local Similarity 72.7%; Pred. No. 0.9;
Matches 72; Conservative 1; Mismatches 26; Indels 0; Gaps 0;

QY 25232 GTGGCTCACCTGTAATCCAGCACTTTGGAGGCCAAGGTAAAGCATCACTTGAGT 25291
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99 GTGACTCACCTATAATCTCTGGCACCTTTGGAGGCTTAGGAAGGAGGATTGTTTGAAC 40

QY 25292 CAGGAGTTAGACACAGCTCTGCCCAACATAGTGAACATC 25330
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 39 CAGGAGCTCAAGACCATCTCTGGAAACATAGCAAGATC 1

RESULT 15
ID T20927
AC T20927 standard; cDNA to mRNA; 103 BP.
DE 24-JUL-1996 (first entry)
KW Human gene signature H0MG02180.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
```

PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1: Page 758-759; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 103 BP; 22 A; 27 C; 21 G; 31 T;  
  
Query Match 0.2%; Score 57.2; DB 1; Length 103;  
Best Local Similarity 77.3%; Pred. No. 0.84;  
Matches 68; Conservative 0; Mismatches 20; Indels 0; Gaps 0;  
  
QY 10254 TCCTCTCCCAAGTAGCTGGGACTACAGTGCATACCCACCGCCCTGCTAATTTTGTGA 10313  
Db 13 TCACCTCCCAAGTAGCTGGGCTAGCTGTGTCGCCACCATGTCCAGCTGATTTNGTA 72  
  
QY 10314 TTTTGTATAGACGGGGTTTCACCATG 10341  
Db 73 TTTTGTATAGACGGGGTTTCACCATG 100  
  
Search completed: June 16, 2000, 12:11:56  
Job time: 170027 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2000, 19:11:36 ; Search time 8514.64 Seconds  
(without alignments)  
13805.343 Million cell updates/sec

Title: US-08-852-495C-1\_COPY\_112000\_141000  
Perfect score: 29001  
Sequence: 1 TTTTCCACTCTTCTTCAG.....AGAGTCTTCACCTCTAGGA 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 156056

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: em\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result	%	Query
SUMMARIES		

No.	Score	Match	Length	DB	ID	Description
1	93.2	0.3	106	37	AA703692	AA703692 ag81a10.r
2	91.4	0.3	109	30	AA243009	AA243009 z125h02.s
3	90.4	0.3	106	30	AA250812	AA250812 z506a05.s
c 4	88.6	0.3	103	84	B48914	B48914 RPI111-4A12
c 5	88.2	0.3	101	39	AA835205	AA835205 ak64h01.s
c 6	88.2	0.3	110	30	AA244245	AA244245 nc07a04.s
c 7	87.8	0.3	101	33	AA381369	AA381369 EST94442
c 8	87.8	0.3	108	84	B65160	B65160 CIT-HSP-201
c 9	86.6	0.3	102	36	AA654562	AA654562 nt75f10.s
c 10	86	0.3	103	108	AQ582186	AQ582186 RPI1-11-4
c 11	86.2	0.3	107	35	AA565533	AA565533 nk42b11.s
c 12	86.4	0.3	107	39	AA828124	AA828124 od71a07.s
c 13	86.6	0.3	110	94	AQ003188	AQ003188 RPI111-1D
c 14	86	0.3	102	36	AA654562	AA654562 nt75f10.s
c 15	85.4	0.3	103	108	AQ535244	AQ535244 RPI1-11-3
c 16	85.2	0.3	109	94	AQ028426	AQ028426 CIT-HSP-2
c 17	85	0.3	110	30	AA244245	AA244245 nc07a04.s
c 18	85.2	0.3	110	39	AA897366	AA897366 am06h02.s
c 19	84.4	0.3	105	21	T94466	T94466 ve35b02.r1
c 20	84.6	0.3	107	103	AQ240182	AQ240182 CIT-HSP-2
c 21	84.2	0.3	106	108	AQ544957	AQ544957 CITBI-EI-
c 22	84.4	0.3	110	106	AQ386882	AQ386882 RPI111-13
c 23	83.8	0.3	103	94	AQ028649	AQ028649 CIT-HSP-2
c 24	83.8	0.3	103	108	AQ535244	AQ535244 RPI1-11-3
c 25	83.6	0.3	106	63	AI991750	AI991750 wt48e01.x
c 26	83.6	0.3	106	63	AI991750	AI991750 wt48e01.x
c 27	83.4	0.3	109	84	B17434	B17434 345K2.IVB C
c 28	83.6	0.3	109	84	B17434	B17434 345K2.IVB C
c 29	83.2	0.3	96	92	AQ936334	AQ936334 RPI1-11-S
c 30	83	0.3	101	33	AA381369	AA381369 EST94442
c 31	83.2	0.3	104	105	AQ321855	AQ321855 RPI111-11
c 32	82.8	0.3	103	38	AA807640	AA807640 nx08b05.s
c 33	82.2	0.3	103	108	AQ584425	AQ584425 RPI1-11-4
c 34	82.2	0.3	105	109	AQ637292	AQ637292 RPI1-11-4
c 35	82.4	0.3	108	84	B32951	B32951 HS-1016-A1-
c 36	81.8	0.3	106	30	AA250812	AA250812 z506a05.s
c 37	82	0.3	106	38	AA812141	AA812141 ob48h02.s
c 38	82	0.3	106	50	AI700000	AI700000 tL36a10.x
c 39	81.6	0.3	104	105	AQ321855	AQ321855 RPI111-11
c 40	81.6	0.3	105	30	AA218889	AA218889 zq15d04.s
c 41	81	0.3	105	28	AA078003	AA078003 7H12D08 C
c 42	81.2	0.3	110	106	AQ386882	AQ386882 RPI111-13
c 43	80.6	0.3	107	33	AA385808	AA385808 EST99495
c 44	80.2	0.3	101	94	AQ076649	AQ076649 CIT-HSP-2
c 45	80.2	0.3	101	105	AQ260734	AQ260734 CITBI-EI-

## ALIGNMENTS

```

RESULT 1
AA703692
LOCUS ag81a10.r1 Striatagene hnt neuron (#937233) Homo sapiens cDNA clone
DEFINITION IMAGE:1140858 5' similar to contains Alu repetitive element; mRNA
sequence.
ACCESSION AA703692
VERSION 106 bp mRNA EST 24-DEC-1997
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 106)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
JOURNAL
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397630.

```

```

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 53.
Location/Qualifiers
source 1. .106
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1140858"
/clone_lib="Striatagene hnt neuron (#937233)"
/lab_host="hnt neurons"
/dev_stage="Striatagene hnt neuron (#937233)"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dT.
Differentiated, post mitotic hnt neurons. Average insert
size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
GAATCGGCACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT 19 a 29 c 29 g 29 t
ORIGIN
Query Match 0.3%; Score 93.2; DB 37; Length 106;
Best Local Similarity 92.5%; Pred. No. 0.2;
Matches 98; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 20240 TTTTGTAGACAGAGGTTTCACCGTGTAGCCGGATGCTCGATCTCCTGACCTCA 20299
|||||
Db 1 TTTTGTAGACAGAGGTTTCACCGTGTAGCCGGATGCTCGATCTCCTGACCTCG 60
|||||
QY 20300 TGATCTGCCCCCTCAGCCTCCCAAGTGTAGGATCAGAGCGATG 20345
|||||
Db 61 TGATCTGCCCCCTCAGCCTCCCAAGTGTAGGATCAGAGCGATG 106
|||||
RESULT 2
AA243009 109 bp mRNA EST 11-MAR-1998
LOCUS z125h02.s1 Striatagene NT2 neuronal precursor 937230 Homo sapiens
DEFINITION cDNA clone IMAGE:66467 3' similar to contains Alu repetitive
element; contains element LTR1 repetitive element ;, mRNA sequence.
ACCESSION AA243009
VERSION 109 bp mRNA EST 11-MAR-1998
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 109)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
JOURNAL
COMMENT On Dec 3, 1996 this sequence version replaced gi:1126869.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 1127 Std Error: 0.00
Seq primer: -4ml3 fwd. ET from Amersham
High quality sequence stop: 102.

```

FEATURES  
source

Location/Qualifiers  
1. .109  
/organism="Homo sapiens"  
/db\_xref="GDB:5426481"  
/db\_xref="taxon:9606"  
/clone="IMAGE:664467"  
/tissue\_type="neuroepithelial cells"  
/dev\_stage="Ntera-2 neuroepithelial cells"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: brain; Vector: phuescript SK-; Site\_1: EcORI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Uninduced, exponentially growing neuroepithelial cells (Ntera-2/cl.D1). Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3"

BASE COUNT 19 a 30 c 30 g 30 t  
ORIGIN

Query Match 0.3%; Score 91.4; DB 30; Length 109;  
Best Local Similarity 89.9%; Pred. No. 0.31; Indels 0; Gaps 0;  
Matches 98; Conservative 0; Mismatches 11

QY 20237 GTATTTTGTAGACACAGGGTTTACCCTGTTAGCGGGATGGTCTCGATCTCCTGACC 20296  
Db 1 GTATTTTGTAGACAGGGGTTTACCCTGTTAGCGGGATGGTCTCGATCTCCTGACC 60

QY 20297 TCATGATCGCCACCTCAGCTCCCAAGTGTAGGATCAGCAGGATG 20345  
Db 61 TCGTGATCGCCCACTCGCGCTCCCAAGTGTGCGGATTACAGGGGTG 109

RESULT 3  
AA250812 106 bp mRNA EST 15-AUG-1997  
LOCUS zs06a05.s1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:684368 3'  
DEFINITION similar to contains Alu repetitive element; contains element MER22  
repetitive element ; mRNA sequence.

ACCESSION AA250812  
VERSION AA250812.1 GI:1885774  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 106)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407356.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 537 Std Error: 0.00  
Seq primer: -41m13 fwd. Et from Amersham  
High quality sequence stop: 87.

FEATURES  
source  
Location/Qualifiers  
1. .106  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:684368"  
/tissue\_type="NCI\_CGAP\_GCB1"  
/lab\_host="DH108"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),  
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman

FEATURES  
source

(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
primed with a Not I - oligo(dT) primer  
[5'-TGTTACCAATCTGAAGTGGAGCGCGCCCTCATTTTTTTTTTTTTTTT-  
3']. Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 20 a 28 c 31 g 27 t  
ORIGIN

Query Match 0.3%; Score 90.4; DB 30; Length 106;  
Best Local Similarity 94.0%; Pred. No. 0.4;  
Matches 94; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 10321 TAGAGACGGGGTTTACCATGTTGGCCAGGCTGCTCAAACTCTTGGAGCTCAGGTGATC 10380  
Db 2 TAGAGACGGGGTTTACCATGTTGGCCAGGCTGCTCAAACTCTTGGAGCTCAGGTGATC 61

QY 10381 CACCTCGCTCGCCCTCCCAAAATGCTGAGATTACAGGTGT 10420  
Db 62 CACTTGCCTTGGCCCTCCCAAAAGTGTGCTGGATTACAGGTGT 101

RESULT 4  
B48914/c 103 bp DNA GSS 08-APR-1999  
LOCUS RPC111-4A12.TP RPCI-11 Homo sapiens genomic clone RPCI-11-4A12,  
DEFINITION genomic survey sequence.  
ACCESSION B48914  
VERSION B48914.1 GI:2601151  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 103)  
AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,  
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and  
Venner,J.C.  
TITLE Use of BAC End Sequences for Sequence-Ready Map Building  
JOURNAL Unpublished (1997)  
COMMENT Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from  
Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/tldb/hungen/bac\_end\_search/bac\_end\_search.html  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source  
Location/Qualifiers  
1. .103  
/organism="Homo sapiens"  
/db\_xref="GDB:7501163"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-4A12"  
/tissue\_type="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACE3.6; Site\_1: EcORI; Site\_2: EcORI;  
RPC111 Human Male BAC Library"

BASE COUNT 30 a 28 c 30 g 15 t  
ORIGIN

Query Match 0.3%; Score 88.6; DB 84; Length 103;  
Best Local Similarity 91.3%; Pred. No. 0.63;  
Matches 94; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 20236 TGTATTTTGTAGAGACAGGGTTTACCGGTGTAGCCGGATGCTCGATCTCTGAC 20295

Db 103 TGTATTTTGTAGAGACAGGGTTTACCGGTGTAGCCGGATGCTCGATCTCTGAC 44

QY 20296 CFCATGATGCCACCTCAGCGTCCCAAGTGCTAGGATCATC 20338

Db 43 CTCGTGATCCGCCCTCGGCTCCCAAGTGCTAGGCTTAC 1

## RESULT 5

LOCUS AA835205 101 bp mRNA EST 23-FEB-1998  
DEFINITION ak64h01.s1 Barstead pancreas HPLRB1 Homo sapiens cDNA clone  
IMAGE:1412689 3' similar to contains Alu repetitive  
element; contains element KER repetitive element ;, mRNA sequence.  
ACCESSION AA835205  
VERSION AA835205.1 GI:2908933  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 101)  
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Stepec, M., Tan, F.,  
Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.  
TITLE WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT On Nov 29, 1993 this sequence version replaced gi:636191.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40ml3 fwd. ET from Amersham.

## FEATURES

Location/Qualifiers  
1..101  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1412689"  
/clone\_lib="Barstead pancreas HPLRB1"  
/sex="female"  
/dev\_stage="adult, 34 years"  
/lab\_host="DH10B"  
/note="Organ: pancreas; Vector: pT7T3D-Pac (Pharmacia)  
with a modified polylinker; Site\_1: EcoRI; Site\_2: NotI;  
1st strand cDNA was primed with a Not I - oligo(dT) primer  
[5'  
TGTTACGAATCTGAAGTGGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTTTTTTTTTT  
3'] ; double-stranded cDNA was ligated to Eco RI adaptors  
[AATCGGATCTTG], digested with Not I and cloned into the  
Not I and Eco RI sites of the modified pT7T3 vector.  
Library constructed by Bob Barstead."  
BASE COUNT 14 a 36 c 27 g 24 t  
ORIGIN

Query Match 0.3%; Score 88.2; DB 39; Length 101;  
Best Local Similarity 92.1%; Pred. No. 0.71;  
Matches 93; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 20081 TGAGAGGAGTCTTCTGTGCGCCAGCTGAATGAGTGGCACAATCTGGGTCACT 20140

Db 1 TGAGAGGAGTCTTCTGTGCGCCAGCTGGAGTGGAGTGGCTGTGATCTCGGCTACT 60

QY 20141 GCAACCTCGGCTCCCGGATTACGCCCATTCCTCTGCTCA 20181

Db 61 GCAAGCTCGGCTCCCGGTTACGCCCATTCCTCTGCTCA 101

## RESULT 6

LOCUS AA244245/c 110 bp mRNA EST 20-AUG-1997  
DEFINITION nc07a04.s1 NCI-CGAP\_Prl Homo sapiens cDNA clone IMAGE:1007406  
similar to contains Alu repetitive element; , mRNA sequence.  
ACCESSION AA244245  
VERSION AA244245.1 GI:1875104  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 110)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 24, 1995 this sequence version replaced gi:634306.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -4ml13 fwd. ET from Amersham  
High quality sequence stop: 90.

## FEATURES

Location/Qualifiers  
1..110  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1007406"  
/clone\_lib="NCI-CGAP\_Prl"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"  
/note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from  
5,000-10,000 microdissected, histologically normal  
prostate epithelial cells. Double-stranded cDNA was  
ligated to EcoRI adaptors, 5 cycles of PCR applied to the  
cDNA with an adaptor-specific primer, and the resulting  
PCR product subcloned into pAMP10 by the UDG-cloning  
method (Life Technologies). Average insert size is 600  
bp. NOTE: Not directionally cloned. This library was  
constructed by David Krizman."  
BASE COUNT 17 a 26 c 28 g 38 t 1 others  
ORIGIN

Query Match 0.3%; Score 88.2; DB 30; Length 110;  
Best Local Similarity 87.3%; Pred. No. 0.68;  
Matches 96; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 13239 TGAGCAGAGATCGCTTGAAACCCAGGAGGAGTTCAGTTCAGTTCCTGTC 13298

Db 110 TGAGCAGAGATCGCTTGAAACCCAGGAGGAGTTCAGTTCAGTTCCTGTC 51

QY 13299 ACTGCACCCCTCGCGGCGAGAGCGAGCTTCGCTCAAAAAACAA 13348

Db 50 ACTGCACCTCAGCTCGCGGCGAGAGCTTCAGTTCAGTTCCTCAAAAAACAA 1

```

RESULT 7
AA381369/c
LOCUS
DEFINITION
  EST94442 Activated T-cells 1 Homo sapiens cDNA 5' end similar to
  EST containing Alu repeat, mRNA sequence.
ACCESSION
  AA381369
VERSION
  AA381369.1 GI:2033689
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 101)
  Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
  Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
  White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
  Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
  Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
  Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
  Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
  Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
  Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
  Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
  Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
  Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
  He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
  Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Melssner,P.S., Olsen,H.,
  Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M.,
  Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
  Fraser,C.M. and Venter,J.C.
  The Institute for Genomic Research
  9712 Medical Center Drive, Rockville, MD 20850 USA
  Tel: 3018699056
  Fax: 3018699423
  Email: arkerlav@tigr.org
  For clone availability, additional sequence and expression
  information related to this EST, please check the TIGR Human Gene
  Index (http://www.tigr.org/tldb/hgi/hgi.html)
  Seq primer: M13 Reverse.
FEATURES
  Location/Qualifiers
  1..101
  /organism="Homo sapiens"
  /db_xref="ATCC (Inhost):185728"
  /db_xref="taxon:9606"
  /clone_lib="Activated T-cells 1"
  /cell_type="T-lymphocyte"
  /dev_stage="adult"
  /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
  XhoI"
BASE COUNT
  18 a 36 c 20 g 25 t 2 others
ORIGIN
  18 a 36 c 20 g 25 t 2 others

Query Match 0.3%; Score 87.8; DB 33; Length 101;
Best Local Similarity 91.1%; Pred. No. 0.78;
Matches 92; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 13186 AAAATTAGCTGGGTGGTGGCGGCACCTGTAATCCACGACTAATCAGGAGCTGAGGCA 13245
Db 101 AAAATTAGCTGGGAGTGGTGGCGGCGCCCTGTATCCAGCTACTCAGGAGGCTGAGGCA 42

QY 13246 GGAGAAATCGGTGAACCCAGGAGGAGAGGTTGCAGTGAGC 13286
Db 41 GGANAATTGCTTGAACCCAGGAGGCGGAGGTTGCAATGAGC 1

RESULT 8
B65160
LOCUS
DEFINITION
  CIT-HSP-2017G2.TRB CIT-HSP Homo sapiens genomic clone 2017G2,
  genomic survey sequence.
ACCESSION
  B65160
VERSION
  B65160.1 GI:2639138
KEYWORDS
  GSS.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 108)
  Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
  Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
  Simon,M. and Venter,J.C.
  Use of a random BAC End Sequence Database for Sequence-Ready Map
  Building
  Unpublished (1997)
  Other_GSSs: CIT-HSP-2017G2.TFB
  Contact: Mark Adams
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: mdadams@tigr.org
  Clones are available from Research Genetics (info@resgen.com). BAC
  end search page:
  http://www.tigr.org/tldb/hungen/bac\_end\_search/bac\_end\_search.html
  Seq primer: M13 Reverse
  Class: BAC ends.
FEATURES
  Location/Qualifiers
  1..108
  /organism="Homo sapiens"
  /db_xref="GDB:7043860"
  /db_xref="taxon:9606"
  /clone="2017G2"
  /clone_lib="CIT-HSP"
  /sex="Male"
  /cell_type="Sperm"
  /note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
  HindIII"
BASE COUNT
  26 a 27 c 34 g 21 t
ORIGIN
  26 a 27 c 34 g 21 t

Query Match 0.3%; Score 87.8; DB 84; Length 108;
Best Local Similarity 88.8%; Pred. No. 0.76;
Matches 95; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 25365 TGGTGGCATGTGCTGTATCCAGCTACTCAGGAGGCTGAGGAGAGTAATGCTTGAA 25424
Db 1 TGGTGGCATGCGCCTGTATCCAGCTACTCAGGAGGCTGAGGAGAGTAATGCTTGAA 60

QY 25425 CCAGAGGCGGAGGTGGCAGTGCAGTGCAGCGGAAATCGCGCACTGCACCTC 25471
Db 61 CCAGGAGGTGGAGGTGGCAGTGCAGTGCAGCGGAAATCATACCTGCACAC 107

RESULT 9
AA654562/c
LOCUS
DEFINITION
  nt75f10.s1 NCI CGAP.Pr3 Homo sapiens cDNA clone IMAGE:1204363
  similar to contains Alu repetitive element; contains element MER22
  repetitive element ;, mRNA sequence.
ACCESSION
  AA654562
VERSION
  AA654562.1 GI:2590716
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens

```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 102)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced gi:1393451.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: David B. Krizman, Ph.D.  
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES  
source

Location/Qualifiers  
1. .102  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1204363"  
/clone\_lib="NCI-CGAP\_Pr3"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"

/note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from  
5,000-10,000 microdissected cells  
histologically determined to be fully malignant prostate  
cancer cells. Double-stranded cDNA was ligated to EcoRI  
adaptors, 5 cycles of PCR applied to the cDNA with an  
adaptor-specific primer, and the resulting PCR product  
subcloned into pAMP10 by the UDP-cloning method (Life  
Technologies). Average insert size is 600 bp. NOTE: Not  
directionally cloned. This library was constructed by  
David Krizman."

BASE COUNT 22 a 32 c 27 g 21 t

ORIGIN

Query Match 0.3%; Score 86.6; DB 36; Length 102;  
Best Local Similarity 91.1%; Pred. No. 1.1;  
Matches 92; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 13071 GCTCATGCTGTAATACACAGCACTTTGGAGGCCGATGTGGTGATCACCTGAGGTCCAG 13130

Db 101 GCTCACTCTGTAATCCAGCACTTTGGAGGCCGATGTGGTGATCACCTGAGGTCCG 42

Qy 13131 GAGTTTGACCACTAGTGGCCCAACATGTTGAACCTCATCT 13171

Db 41 GAGTTTGACCACTAGTGGCCCAACATGTTGAACCTCATCT 1

RESULT 10

AQ582186

LOCUS

DEFINITION AQ582186 103 bp DNA GSS 07-JUN-1999

RPCL11-451A15.TJ RPCI-11 Homo sapiens genomic clone

RPCL11-451A15, genomic survey sequence.

ACCESSION AQ582186

VERSION AQ582186.1 GI:5009296

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 103)

REFERENCE Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and

TITLE

JOURNAL

COMMENT

Venter, J.C.  
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
Map Building  
Unpublished (1997)  
On Feb 19, 1999 this sequence version replaced gi:4146076.  
Other GSSs: RPCI-11-451A15.TJ  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@tigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from  
Research Genet cs (<http://info@resgen.com>). BAC end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html).  
Seq primer: Sp6  
Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
1. .103  
/organism="Homo sapiens"  
/db\_xref="GDB:7672814"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-451A15"  
/clone\_lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPCI11 Human Male BAC Library"

BASE COUNT 19 a 36 c 25 g 22 t  
ORIGIN 1 others

Query Match 0.3%; Score 86; DB 108; Length 103;  
Best Local Similarity 89.3%; Pred. No. 1.2;  
Matches 92; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 20112 GGAATGCGAGTGGCACAATCTCGGCTCACTGCACCTCGCTCCCGGATTCAGGCATTC 20171  
Db 1 GGAGTGCNGTGGCACAATCTCGGCTCACTGCACCTCGCTCCCGGATTCAGGCATTC 60

Qy 20172 TCCTGCCTCAACCTCCCGAGTAGTGGGACCAACGCGCCGC 20214

Db 61 TCCTGCCTCAGCTCCCGAGTAGTGGGACTACAGCGCTGC 103

RESULT 11

AA565533/c

LOCUS

DEFINITION AA565533 107 bp mRNA EST 08-SEP-1997

nk42b11.s1 NCI-CGAP\_GC2 Homo sapiens cDNA clone IMAGE:1016157 3'

similar to contains Alu repetitive element; mRNA sequence.

ACCESSION AA565533

VERSION AA565533.1 GI:2337172

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 107)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

On Sep 12, 1996 this sequence version replaced gi:1393355.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Stratagene, Inc., David B. Krizman,



```
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"
BASE COUNT      22 a  27 c  26 g  35 t
ORIGIN

Query Match      0.3%; Score 86.6; DB 94; Length 110;
Best Local Similarity 87.2%; Pred. No. 1;
Matches 95; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 20232 TTTTGTATTTTAGTAGACAGAGGTTTCACCGTGTAGCGGGATGGTCGATCTCC 20291
|||||
Db 2 TTTTGTATTTTAGTAGACAGGTTTACCATGTGTGGCCAGGATGTCGCCGATCTCT 61
|||||

QY 20292 TGACCTCATGATCTGCCACCTCAGCCCTCCCAAGTGTAGGATCACAG 20340
|||||
Db 62 TGACCTCATGATCCACCTGCCCGCCAGCTCCCAAGTGTGGGATTACAG 110
|||||

RESULT 14
AA654562
LOCUS      AA654562      102 bp      mRNA      EST      04-NOV-1997
DEFINITION nt75f10.s1 NCI_CGAP_Pr3 Homo sapiens cDNA clone IMAGE:1204363
similar to contains Alu repetitive element;contains element MER22
repetitive element ;, mRNA sequence.
ACCESSION  AA654562
VERSION     AA654562.1 GI:2590716
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 102)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1393451.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
M.D., Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.W.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source           1..102
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /clone="IMAGE:1204363"
     /clone_lib="NCI_CGAP_Pr3"
     /sex="Male"
     /dev_stage="45 years old"
     /lab_host="DH10B"
     /note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected cells
histologically-determined to be fully malignant prostate
cancer cells. Double-stranded cDNA was ligated to EcoRI
adaptors, 5 cycles of PCR applied to the cDNA with an
adaptor-specific primer, and the resulting PCR product
subcloned into pAMP10 by the UDG-cloning method (Life
Technologies). Average insert size is 600 bp. NOTE: Not
directionally cloned. This library was constructed by
David Krizman."
BASE COUNT      22 a  32 c  27 g  21 t
```

```
ORIGIN

Query Match      0.3%; Score 86; DB 36; Length 102;
Best Local Similarity 90.2%; Pred. No. 1.2;
Matches 92; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 10324 AGACGGGGTTTCACCATGTGGCCAGGTGGTCTCAAACTCCTGACCTCAGGTGATCCAC 10383
|||||
Db 1 AGACAGGGTTTCACCATGTGGCCAGTCTGTCTCAAACTCCGGACCTCAGGTAATCCGC 60
|||||

QY 10384 CTGCTCGCGCTCCCAAAATGCTGAGATTACAGGTGTGAGCC 10425
|||||
Db 61 CCACCTCGCGCTCCCAAGTGTGGGATTACAGGAGTGAGCC 102
|||||

RESULT 15
AQ535244/c
LOCUS      AQ535244      103 bp      DNA      GSS      18-MAY-1999
DEFINITION RPCI-11-317H22.TV RPCI-11 Homo sapiens genomic clone
RPCI-11-317H22, genomic survey sequence.
ACCESSION  AQ535244
VERSION     AQ535244.1 GI:4846934
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 103)
AUTHORS     Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
Venter,J.C.
TITLE       Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL     Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
The Institute of Eukaryotic Genomics
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pleter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: 17
Class: BAC ends.
FEATURES             Location/Qualifiers
     source           1..103
     /organism="Homo sapiens"
     /db_xref="GDB:7621533"
     /db_xref="taxon:9606"
     /clone="RPCI-11-317H22"
     /clone_lib="RPCI-11"
     /sex="Male"
     /cell_type="Lymphocytes"
     /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"
BASE COUNT      31 a  27 c  27 g  18 t
ORIGIN

Query Match      0.3%; Score 85.4; DB 108; Length 103;
Best Local Similarity 89.3%; Pred. No. 1.4;
Matches 92; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 7522 TTTTGTATTTTAGCAGAGATGGGTTTCACCATGTGGCCAGACTGGTCFCAACTCC 7581
|||||
Db 103 TTTTGTATTTATATACACAGAGCGGGGTTTCACCATGTGGCCAGGCTGGTCTCGAATCC 44
|||||

QY 7582 TGGCCTCAAGTGATGCTGGCCACCTCGGCTCCCGAAGTGCTGG 7624
|||||
```



Db 43 TGACCTCAAGTCATCTGCCCGTCTTGGCCTCCCAAAGTCTGG 1

Search completed: June 16, 2000, 03:46:29  
Job time: 140956 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2000, 00:17:42 ; Search time 372.34 Seconds  
(without alignments)  
10124.339 Million cell updates/sec

Title: US-08-852-495C-1\_COPY\_112000\_141000  
Perfect score: 29001  
Sequence: 1 TTTTCCACTCTTCTTCAG.....AGACTGTTGACCTCTAGGA 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 374504

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/6\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/PCUS\_COMB.seq:\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78.4	0.3	105	4 US-08-481-658B-65	Sequence 65, Appl
2	78.4	0.3	105	4 US-08-477-504A-65	Sequence 65, Appl
3	78.4	0.3	105	4 US-08-486-756A-65	Sequence 65, Appl
4	78.4	0.3	105	4 US-08-485-862B-65	Sequence 65, Appl
5	78.4	0.3	105	5 US-08-787-739-65	Sequence 65, Appl
6	68	0.2	105	4 US-08-481-658B-65	Sequence 65, Appl
7	68	0.2	105	4 US-08-477-504A-65	Sequence 65, Appl
8	68	0.2	105	4 US-08-486-756A-65	Sequence 65, Appl
9	68	0.2	105	4 US-08-485-862B-65	Sequence 65, Appl
10	68	0.2	105	5 US-08-787-739-65	Sequence 65, Appl
11	60.4	0.2	78	3 US-08-454-557C-70	Sequence 70, Appl
12	60.4	0.2	78	4 US-08-340-426D-70	Sequence 70, Appl
13	60.4	0.2	78	4 US-08-450-673C-70	Sequence 70, Appl
14	60.4	0.2	78	6 PCT-US95-17111A-70	Sequence 70, Appl
15	58.8	0.2	78	3 US-08-454-557C-70	Sequence 70, Appl
16	58.8	0.2	78	4 US-08-340-426D-70	Sequence 70, Appl
17	58.8	0.2	78	4 US-08-450-673C-70	Sequence 70, Appl
18	58.8	0.2	78	6 PCT-US95-17111A-70	Sequence 70, Appl
19	59	0.2	84	3 US-08-454-557C-91	Sequence 91, Appl
20	59	0.2	84	4 US-08-340-426D-91	Sequence 91, Appl
21	59	0.2	84	4 US-08-450-673C-91	Sequence 91, Appl
22	59	0.2	84	6 PCT-US95-17111A-91	Sequence 91, Appl
23	53.2	0.2	76	3 US-08-454-557C-69	Sequence 69, Appl
24	53.2	0.2	76	4 US-08-340-426D-69	Sequence 69, Appl
25	53.2	0.2	76	4 US-08-450-673C-69	Sequence 69, Appl
26	53.2	0.2	76	6 PCT-US95-17111A-69	Sequence 69, Appl
27	53.4	0.2	85	3 US-08-454-557C-92	Sequence 92, Appl

Sequence 92, Appl  
Sequence 92, Appl  
Sequence 92, Appl  
Sequence 91, Appl  
Sequence 91, Appl  
Sequence 91, Appl  
Sequence 57, Appl  
Sequence 57, Appl  
Sequence 57, Appl  
Sequence 57, Appl  
Sequence 57, Appl  
Sequence 69, Appl  
Sequence 69, Appl  
Sequence 69, Appl  
Sequence 69, Appl  
Sequence 66, Appl  
Sequence 66, Appl  
Sequence 66, Appl

ALIGNMENTS

RESULT 1  
US-08-481-658B-65  
; Sequence 65, Application US/08481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,658B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauger, Leona L.  
; REFERENCE/DOCKET NUMBER: D-0021.3E  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-0727  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; US-08-481-658B-65

Query Match 0.3%; Score 78.4; DB 4; Length 105;  
Best Local Similarity 84.6%; Pred. No. 2.8e-08;

Matches 88; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 19437 TTTTGTATTTTACTAAGAGGGGTTTCACCATGTTGGTCAGGCTGCTCTCCAACTCC 19496

Db 2 TTTTACATCTTTAGTAGAGCAGGGTTTCCACCATTTGGCCAGGCTGCTCTCAAACTCC 61

QY 19497 TGACCTCATGATCTGCCACCTTGGCCCTCCCAAAAGTGTGGGAT 19540

Db 62 TGACCTTGTGATCCACGAGCTCGGCCCTCCCAAAAGTGTGGGAT 105

## RESULT 2

US-08-477-504A-65

; Sequence 65, Application US/08477504A

; Patent No. 5972353

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

; CITY: Tiburon

; STATE: California

; COUNTRY: USA

; ZIP: 94920

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/477,504A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/260,190

; FILING DATE: 15-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lauder, Leona L.

; REGISTRATION NUMBER: 30,863

; REFERENCE/DOCKET NUMBER: D-0021.3D

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-435-2034

; TELEFAX: 415-435-0727

; INFORMATION FOR SEQ ID NO: 65:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 105 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-477-504A-65

Query Match 0.3%; Score 78.4; DB 4; Length 105;

Best Local Similarity 84.6%; Pred. No. 2.8e-08;

Matches 88; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 19437 TTTTGTATTTTACTAAGAGGGGTTTCACCATGTTGGTCAGGCTGCTCTCCAACTCC 19496

Db 2 TTTTACATCTTTAGTAGAGCAGGGTTTCCACCATTTGGCCAGGCTGCTCTCAAACTCC 61

QY 19497 TGACCTCATGATCTGCCACCTTGGCCCTCCCAAAAGTGTGGGAT 19540

Db 62 TGACCTTGTGATCCACGAGCTCGGCCCTCCCAAAAGTGTGGGAT 105

## RESULT 3

US-08-486-756A-65

; Sequence 65, Application US/08486756A

; Patent No. 5981711

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

; CITY: Tiburon

; STATE: California

; COUNTRY: USA

; ZIP: 94920

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/486,756A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/260,190

; FILING DATE: 15-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Lauder, Leona L.

; REGISTRATION NUMBER: 30,863

; REFERENCE/DOCKET NUMBER: D-0021.3C

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-435-2034

; TELEFAX: 415-435-0727

; INFORMATION FOR SEQ ID NO: 65:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 105 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-486-756A-65

Query Match 0.3%; Score 78.4; DB 4; Length 105;

Best Local Similarity 84.6%; Pred. No. 2.8e-08;

Matches 88; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 19437 TTTTGTATTTTACTAAGAGGGGTTTCACCATGTTGGTCAGGCTGCTCTCCAACTCC 19496

Db 2 TTTTACATCTTTAGTAGAGCAGGGTTTCCACCATTTGGCCAGGCTGCTCTCAAACTCC 61

QY 19497 TGACCTCATGATCTGCCACCTTGGCCCTCCCAAAAGTGTGGGAT 19540

Db 62 TGACCTTGTGATCCACGAGCTCGGCCCTCCCAAAAGTGTGGGAT 105

## RESULT 4

US-08-485-862B-65

; Sequence 65, Application US/08485862B

; Patent No. 5989838

; GENERAL INFORMATION:

; APPLICANT: Zavada, Jan

; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir

; TITLE OF INVENTION: MN Gene and Protein

; NUMBER OF SEQUENCES: 86

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Leona L. Lauder

; STREET: 6 Mariposa Court

```

; APPLICATION NUMBER: US/08/787,739
; FILING DATE: 24-JAN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,049
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/486,756
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/477,504
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/481,658
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,862
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,863
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,077
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0021.4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-981-2034
; TELEFAX: 415-981-0332
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-787-739-65

Query Match 0.3%; Score 78.4; DB 5; Length 105;
Best Local Similarity 84.6%; Pred. No. 2.8e-08;
Matches 88; Conservative 0; Mismatches 16; Indels 0; Gaps

QY 19437 TTTTGTATTTTGTAGTAAGACGGGGTTTCACCATGTTGGTCAGCGTGTCTCAACTCC 19490
      |||||  ||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 2 TTTTATCATCTTTAGTAGACAGGGTTTCACCATATTGGCAGCGTGTCTCAACTCC 61

QY 19497 TGACCTCATGATCGCCACCTTCGGCTCCCAAGTGTCTGGAT 19540
      |||||  |||||  ||  ||  |||||  |||||  |||||  |||||  |||||  |||||
Db 62 TGACCTGTGATCACCAGCGCTCGGCCTCCCAAGTGTCTGGAT 105

RESULT 6
US-08-481-658B-65/c
; Sequence 65, Application US/08481658B
; Patent No. 5955075
; GENERAL INFORMATION:
; APPLICANT: Zavada, Jan
; APPLICANT: Pastorekova, Silvia
; APPLICANT: Pastorek, Jaromir
; TITLE OF INVENTION: MN Gene and Protein
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/481.658B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-481-658B-65

Query Match 0.2%; Score 68; DB 4; Length 105;  
Best Local Similarity 84.0%; Pred. No. 5e-06;  
Matches 89; Conservative 0; Mismatches 15; Indels 2; Gaps 1;  
QY 13084 ATACCAGCACTTTGGAGGCCGATGGGTGATCACCTGAGTCAGGTCAGGAGTTTGAGACCA 13143  
DB 105 ATCCAGCACTTTGGAGGCCGAGGCTGGTGATCAC--AAGGTCAGGAGTTTGAGAGCA 48  
QY 13144 GACTGCCCAACATGTGGAACCTCATCTCTAGTAAAAATACAAAA 13189  
DB 47 GCCTGCCCAATATGTGGAACCCCTGCTCTACTAAGATGTAAAAA 2

RESULT 7  
US-08-477-504A-65/c  
Sequence 65, Application US/08477504A  
Patent No. 5972353  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477.504A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:

NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-477-504A-65  
Query Match 0.2%; Score 68; DB 4; Length 105;  
Best Local Similarity 84.0%; Pred. No. 5e-06;  
Matches 89; Conservative 0; Mismatches 15; Indels 2; Gaps 1;  
QY 13084 ATACCAGCACTTTGGAGGCCGATGGGTGATCACCTGAGTCAGGTCAGGAGTTTGAGACCA 13143  
DB 105 ATCCAGCACTTTGGAGGCCGAGGCTGGTGATCAC--AAGGTCAGGAGTTTGAGAGCA 48  
QY 13144 GACTGCCCAACATGTGGAACCTCATCTCTAGTAAAAATACAAAA 13189  
DB 47 GCCTGCCCAATATGTGGAACCCCTGCTCTACTAAGATGTAAAAA 2

RESULT 8  
US-08-486-756A-65/c  
Sequence 65, Application US/08486756A  
Patent No. 5981711  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486.756A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-486-756A-65

Query Match 0.2%; Score 68; DB 4; Length 105;  
Best Local Similarity 84.0%; Pred. No. 5e-06;  
Matches 89; Conservative 0; Mismatches 15; Indels 2; Gaps 1;

QY 13084 ATACGAGCACTTTGGAGGCCGATGTTGGTGGATCAGTCAGGAGTTTGAGACCA 13143  
|| |||||  
DB 105 ATCCGAGCACTTTGGAGGCCGAGGCTGGTGATCAC--AAGGTCAGGAGTTTGAGAGCA 48  
|| |||||

QY 13144 GACTGCCCAACATGGTGAACCTCATCTCTAGTAAAAATACAAAA 13189  
|| |||||  
DB 47 GCCTGCCAATATGGTGAACCTCTCTACTAAAGATGTAATAA 2  
|| |||||

RESULT 9  
US-08-485-862B-65/c  
; Sequence 65, Application US/08485862B  
; Patent No. 5989838  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION NUMBER: US/08/485,862B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/477,504  
; FILING DATE: 07-JUN-1995  
; APPLICATION NUMBER: US 08/260,190  
; FILING DATE: 15-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.3D  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-435-2034  
; TELEFAX: 415-435-0727  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-485-862B-65

Query Match 0.2%; Score 68; DB 4; Length 105;  
Best Local Similarity 84.0%; Pred. No. 5e-06;  
Matches 89; Conservative 0; Mismatches 15; Indels 2; Gaps 1;

QY 13084 ATACGAGCACTTTGGAGGCCGATGTTGGTGGATCAGTCAGGAGTTTGAGACCA 13143  
|| |||||  
DB 105 ATCCGAGCACTTTGGAGGCCGAGGCTGGTGATCAC--AAGGTCAGGAGTTTGAGAGCA 48  
|| |||||  
QY 13144 GACTGCCCAACATGGTGAACCTCATCTCTAGTAAAAATACAAAA 13189  
|| |||||  
DB 47 GCCTGCCAATATGGTGAACCTCTCTACTAAAGATGTAATAA 2  
|| |||||

RESULT 10  
US-08-787-739-65/c  
; Sequence 65, Application US/08787739  
; Patent No. 6027887  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia  
; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 96  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 369 Pine Street, Suite 610  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION NUMBER: US/08/787,739  
; FILING DATE: 24-JAN-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,049  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/486,756  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/477,504  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/481,658  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,862  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,863  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/487,077  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lauder, Leona L.  
; REGISTRATION NUMBER: 30,863  
; REFERENCE/DOCKET NUMBER: D-0021.4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-981-2034  
; TELEFAX: 415-981-0332  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 105 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-787-739-65

```
Query Match          0.2%; Score 68; DB 5; Length 105;
Best Local Similarity 84.0%; Pred. No. 5e-06;
Matches 89; Conservative 0; Mismatches 15; Indels 2; Gaps 1;

QY 13084 ATACAGCACTTTGGGAGCCGATGTGGTGATCCTGATGAGTGGAGGAGTTGAGACCA 13143
      |||||||
Db 105 ATCCAGCACTTTGGGAGCCGAGGCTGTGGATCAC--AAGTCAAGGAGTTGAGAGCA 48
      |||||||

QY 13144 GACTGCCAACATGGTGAACCTCATCTCTAGTAAATACAAAAA 13189
      |||||||
Db 47 GCCTGCCAATATGGTGAACCCCTGTCTACTAAAGATGTAAAAA 2

RESULT 11
US-08-454-557C-70
; Sequence 70, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,557C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; US-08-454-557C-70

Query Match          0.2%; Score 60.4; DB 3; Length 78;
Best Local Similarity 85.9%; Pred. No. 0.00019;
Matches 67; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 19422 ACCAAGCCCGGCTAATTTTGTATTTTAAAGACGGGTTTCACCATGTTGGTCAGG 19481
      || |||||
Db 1 ACAAGCCCGCAGCTAATTTGTATTTTAGTAGAGATGGGTTTCTCCATGTTTCATCAGG 60

QY 19482 CTGGTCTCCAACCTCTGA 19499
      |||||
Db 61 CTGGTCTCGAACTCCTGA 78

RESULT 13
US-08-450-673C-70
; Sequence 70, Application US/08450673C
; Patent No. 5948888
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

Query Match          0.2%; Score 60.4; DB 3; Length 78;
Best Local Similarity 85.9%; Pred. No. 0.00019;
Matches 67; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 19422 ACCAAGCCCGGCTAATTTTGTATTTTAAAGACGGGTTTCACCATGTTGGTCAGG 19481
      || |||||
Db 1 ACAAGCCCGCAGCTAATTTGTATTTTAGTAGAGATGGGTTTCTCCATGTTTCATCAGG 60

QY 19482 CTGGTCTCCAACCTCTGA 19499
      |||||
Db 61 CTGGTCTCGAACTCCTGA 78

RESULT 12
US-08-340-426D-70
; Sequence 70, Application US/08340426D
; Patent No. 5948634
```



```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450.673C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
;
US-08-450-673C-70

Query Match 0.2%; Score 60.4; DB 4; Length 78;
Best Local Similarity 85.9%; Pred. No. 0.00019;
Matches 67; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 19422 ACCAAGCCCGGCTAATTTTGTATTTTAGTAAGACGGGTTTCACCATGTTGGTCAGG 19481
Db 1 ACAAGCCCGCAGCTAATTTGTATTTTAGTAGAGATGGGGTTCTCCATGTTTCATCAGG 60

QY 19482 CTGGTCTCCAACCTCTGA 19499
Db 61 CTGGGTGCGAACTCCTGA 78

RESULT 14
PCT-US95-17111A-70
; Sequence 70, Application PC/TUS951711A
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and
; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
;
US-08-454-557C-70

Query Match 0.2%; Score 58.8; DB 3; Length 78;
Best Local Similarity 84.6%; Pred. No. 0.00042;
Matches 66; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 13127 TCAGGAGTTTGACACGACACTGCCCAACATGGTGAACCTCATCTCTATAAATAACAA 13186
Db 78 TCAGGAGTTTCGACACCGACGCTGATGAACATGGAGAACCCCATCTCTACTAAAAATAACAA 19

QY 13187 AAATTAGCTGGGTGGGT 13204
Db 18 ATATTAGCTGGGGGTGT 1

```

```

;
; TOPOLOGY: both
;
PCT-US95-17111A-70

Query Match 0.2%; Score 60.4; DB 6; Length 78;
Best Local Similarity 85.9%; Pred. No. 0.00019;
Matches 67; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 19422 ACCAAGCCCGGCTAATTTTGTATTTTAGTAAGACGGGTTTCACCATGTTGGTCAGG 19481
Db 1 ACAAGCCCGCAGCTAATTTGTATTTTAGTAGAGATGGGGTTCTCCATGTTTCATCAGG 60

QY 19482 CTGGTCTCCAACCTCTGA 19499
Db 61 CTGGGTGCGAACTCCTGA 78

RESULT 15
US-08-454-557C-70/C
; Sequence 70, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,557C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
;
US-08-454-557C-70

Query Match 0.2%; Score 58.8; DB 3; Length 78;
Best Local Similarity 84.6%; Pred. No. 0.00042;
Matches 66; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 13127 TCAGGAGTTTGACACGACACTGCCCAACATGGTGAACCTCATCTCTATAAATAACAA 13186
Db 78 TCAGGAGTTTCGACACCGACGCTGATGAACATGGAGAACCCCATCTCTACTAAAAATAACAA 19

QY 13187 AAATTAGCTGGGTGGGT 13204
Db 18 ATATTAGCTGGGGGTGT 1

```

Search completed: June 16, 2000, 08:52:43  
Job time: 158293 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2000, 08:46:10 ; Search time 17973.3 Seconds  
(without alignments)  
-1569.657 Million cell updates/sec

Title: US-08-852-495C-1\_COPY\_140000\_169000  
Perfect score: 29001  
Sequence: 1 CCCTCAATCCATATGCAC.....TGTATACACAGAAAGTTACC 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 370290

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pl1.\*  
8: gb\_pl2.\*  
9: gb\_pr1.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: gb\_ro.\*  
13: gb\_sts.\*  
14: gb\_sy.\*  
15: gb\_un.\*  
16: gb\_vi.\*  
17: em\_fun.\*  
18: em\_hum1.\*  
19: em\_hum2.\*  
20: em\_in.\*  
21: em\_om.\*  
22: em\_or.\*  
23: em\_ov.\*  
24: em\_pat.\*  
25: em\_ph.\*  
26: em\_pl.\*  
27: em\_ro.\*  
28: em\_sts.\*  
29: em\_sy.\*  
30: em\_un.\*  
31: em\_vi.\*  
32: gb\_htg1.\*  
33: gb\_htg2.\*  
34: gb\_in1.\*  
35: gb\_in2.\*  
36: em\_ba1.\*  
37: em\_ba2.\*  
38: em\_hum3.\*  
39: em\_hum4.\*  
40: gb\_pr4.\*  
41: gb\_htg3.\*  
42: gb\_htg4.\*  
43: gb\_htg5.\*  
44: gb\_htg6.\*

45: gb\_htg7.\*  
46: em\_htg1.\*  
47: em\_htg2.\*  
48: em\_htg3.\*  
49: em\_hum5.\*  
50: gb\_pl3.\*  
51: gb\_pr5.\*  
52: gb\_htg8.\*  
53: gb\_htg9.\*  
54: gb\_htg10.\*  
55: gb\_htg11.\*  
56: gb\_htg12.\*  
57: gb\_htg13.\*  
58: gb\_htg14.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	85.8	0.3	108	11 HSU67803	U67803 Human small
2	85.2	0.3	108	10 HSLDLRN2	X05250 Human LDL-r
3	80.6	0.3	107	9 HUMALCE162	M87924 Human carc1
4	79.2	0.3	104	9 HUMALCE272	M87899 Human carc1
5	77.6	0.3	108	10 HSLDLRN2	X05250 Human LDL-r
6	77.8	0.3	108	11 HSU67804	U67804 Human small
7	76	0.3	103	9 HUMALCE221	M87896 Human carc1
8	76.2	0.3	108	11 HSU67808	U67808 Human small
9	75.4	0.3	103	13 HS8IC8R	X57789 Human sequ
10	75	0.3	97	9 HUMLDLRA1	M14178 Human low d
11	75	0.3	107	9 HUMALCE162	M87924 Human carc1
12	74.4	0.3	103	13 HS8IC8R	X57789 Human sequ
13	74.6	0.3	108	10 HSLDLRD1	X05249 Human LDL-r
14	74.6	0.3	108	10 HSLDLRD2	X05251 Human LDL-r
15	74.4	0.3	110	11 HSU67807	U67807 Human small
16	74	0.3	91	13 HUMUT8164A	L30244 Human STS U
17	72.4	0.2	101	10 S79560	S79560 HRX (intron
18	72.4	0.2	103	9 HUMALCE221	M87896 Human carc1
19	72.4	0.2	108	10 HSLDLI12	X05248 Human LDL-r
20	72.6	0.3	108	13 G43535	G43535 WIAF-2393-S
21	72.4	0.2	110	9 HUMALCE43	M87900 Human carc1
22	72	0.2	97	9 HUMLDLRA2	M14180 Human low d
23	71.8	0.2	110	11 HSU67807	U67807 Human small
24	71.4	0.2	97	9 HUMLDLDRJ	M14179 Human famil
25	71	0.2	97	9 HUMLDLRA2	M14180 Human low d
26	70.8	0.2	100	13 HUMUT931A	L31299 Human STS U
27	71	0.2	105	13 G32655	G32655 A009L30 Hum
28	70.2	0.2	97	9 HUMLDLRA1	M14178 Human low d
29	69.8	0.2	107	11 HSU67806	U67806 Human small
30	69.4	0.2	108	9 HUMDID03M5	D16965 Human HepG2
31	68.4	0.2	95	13 HUMUT8002B	L30176 Human STS U
32	68.4	0.2	95	13 HUMUT8002B	L30176 Human STS U
33	68.6	-0.2	97	9 HUMLDLDRJ	M14179 Human famil
34	68.6	0.2	108	10 HSLDLRD1	X05249 Human LDL-r
35	68.6	0.2	108	10 HSLDLRD2	X05251 Human LDL-r
36	67.6	0.2	90	9 HUMLDLDRM	M15365 Human low d
37	67.8	0.2	108	13 G43535	G43535 WIAF-2393-S
38	67.2	0.2	100	9 HUMGALNSA	D45223 Human GALNS
39	66.8	0.2	95	10 HSSTHPKIB	X66361 H.sapiens m
40	66	0.2	100	10 HSLAS27	X91545 H.sapiens D
41	66	0.2	100	13 HUMUT931A	L31299 Human STS U
42	65.6	0.2	90	9 HUMLDLDRFL	K03555 Human low d
43	65.8	0.2	101	10 S79560	S79560 HRX (intron
44	65.4	0.2	84	5 AR051521	AR051521 Sequence
45	65.6	0.2	102	13 G32906	G32906 A009W09 Hum

ALIGNMENTS

```
RESULT 1
LOCUS HSU67803 108 bp RNA PRI 01-AUG-1997
DEFINITION Human small cytoplasmic Alu transcript.
ACCESSION U67803
VERSION U67803.1 GI:2289917
KEYWORDS Alu.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE cDNAs derived from primary and small cytoplasmic Alu (scAlu)
transcripts
JOURNAL J. Mol. Biol. 271 (2), 222-234 (1997)
MEDLINE 97415756
REFERENCE 2 (bases 1 to 108)
AUTHORS Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The
Children's Hospital of Philadelphia, 1004F Abramson Research
Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA
LOCATION/Qualifiers
FEATURES
source
1..108
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TscAlu2"
1..108
/repeat_region
1..108
/note="scAlu"
/rpt_family="Alu"
/rpt_type="dispersed"
BASE COUNT 23 a 39 c 30 g 16 t
ORIGIN
Query Match 0.3%; Score 85.8; DB 11; Length 108;
Best Local Similarity 92.8%; Pred. No. 0.00045;
Matches 90; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 24035 GCCTGTATCCAGCACATTGGAGCGTGAGTGGTGAATCACAGAGTCAGGAGATCAA 24094
|||||
Db 1 GCCTGTATCCAGCACATTGGAGCGCGGCGGCGGATCACAGGTCAGGAGATCGA 60
|||||

QY 24095 GACCATCTGGCCACATGTTGAACCCCGTCTCTAC 24131
|||||
Db 61 GACCATCTGGCTAACAAGGTGAACCCCGTCTCTAC 97
|||||

RESULT 2
LOCUS HSLDLRN2 108 bp DNA PRI 20-MAY-1992
DEFINITION Human LDL-receptor gene intron 14 fragment (normal gene).
ACCESSION X05250
VERSION X05250.1 GI:34337
KEYWORDS Alu repetitive sequence; low density lipoprotein receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R.,
Williamson,R. and Humphries,S.
TITLE Unequal crossing-over between two alu-repetitive DNA sequences in
the low-density-lipoprotein-receptor gene. A possible mechanism for
the defect in a patient with familial hypercholesterolaemia
JOURNAL Eur. J. Biochem. 164 (1), 77-81 (1987)
MEDLINE 87161901
COMMENT See X05252 for deletion junction
DATA kindly reviewed (07-DEC-1987) by HUMPHRIES S.
LOCATION/Qualifiers
1..108
FEATURES
source
```

```
intron
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..108
/note="intron XIV fragment"
BASE COUNT 28 a 23 c 39 g 18 t
ORIGIN
Query Match 0.3%; Score 85.2; DB 10; Length 108;
Best Local Similarity 87.7%; Pred. No. 0.00055;
Matches 93; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 24141 AAAAATTAGCCAGGCGATGCTAGCACATCCCTGTATATCCAGCTACTCAAGAGCGTGAGGC 24200
|||||
Db 3 AAAAATTAGCCAGGCGGTGGTGGCAGGTGCTGTATATCCAGCTACTCGGAGGCTGAGGC 62
|||||

QY 24201 AGGGAATTGTTGAACCCGGGAGGTGGACATTCAGTGTAGCTGAG 24246
|||||
Db 63 AGGAGAATTGCTTGAACCCAGGAGGCAGAGGTGCGAGTGAGCCGAG 108
|||||

RESULT 3
LOCUS HUMALCE162 107 bp ss-RNA PRI 15-APR-1994
DEFINITION Human carcinoma cell-derived Alu RNA transcript, clone CE162.
ACCESSION M87924
VERSION M87924.1 GI:174871
KEYWORDS Alu repeat.
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 107)
AUTHORS Sinnett,D., Richer,C., Deragon,J.-M. and Labuda,D.
TITLE Alu RNA transcripts in human embryonal carcinoma cells. Model of
post-transcriptional selection of master sequences
JOURNAL J. Mol. Biol. (1992) In press
FEATURES
Location/Qualifiers
1..107
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="NTERa2D1"
/dev_stage="embryo"
/sex="male"
/tissue_type="carcinoma"
BASE COUNT 28 a 30 c 35 g 14 t
ORIGIN
Query Match 0.3%; Score 80.6; DB 9; Length 107;
Best Local Similarity 86.4%; Pred. No. 0.0028;
Matches 89; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 729 GCGAGGAGAATCACTTGAACCTGGGAGGCAGAGGTTGAGTGAGGGAGATGGCGCCACT 788
|||||
Db 5 GCGAGAAGAATGGGTGAACCCGGGAGCGGAGCTTGCAGTGAGCGAGATCGCGCCACT 64
|||||

QY 789 GCATCCAGCTGACCAACACAGCGAGACTCTGTCTCAAAAAA 831
|||||
Db 65 GCATCCAGCGCTGGCGGACAGAGCGGAGACTCCGTCTCAAAAAA 107
|||||

RESULT 4
LOCUS HUMALCE272 104 bp ss-RNA PRI 15-APR-1994
DEFINITION Human carcinoma cell-derived Alu RNA transcript, clone CE272.
ACCESSION M87899
VERSION M87899.1 GI:174875
KEYWORDS Alu repeat.
SOURCE Homo sapiens male embryo carcinoma cDNA to other RNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

[illegible]

RESULT	6	HSU67804	108 bp	RNA	PRI	01-AUG-1997
LOCUS		Human small cytoplasmic Alu transcript.				
DEFINITION						
ACCESSION		U67804				
VERSION		U67804.1	GI:2289918			
KEYWORDS		Alu.				
SOURCE		human.				
ORGANISM		Homo sapiens				
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE		1 (bases 1 to 108)				
AUTHORS		Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.				
TITLE		cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts				
JOURNAL		J. Mol. Biol. 271 (2), 222-234 (1997)				
MEDLINE		97415756				
REFERENCE		2 (bases 1 to 108)				
AUTHORS		Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.				
TITLE		Direct Submission				
JOURNAL		Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The Children's Hospital of Philadelphia, 1004F Abramson Research Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA				
FEATURES		Location/Qualifiers				
source		1..108				
		/organism="Homo sapiens"				
		/db_xref="taxon:9606"				
		/clone="TscAlu3"				
repeat_region		1..108				
		/note="scAlu"				
		/rpt_family="Alu"				
		/rpt_type="dispersed"				
BASE COUNT		26 a 38 c 26 g				
ORIGIN		18 t				
Query Match		0.3%; Score 77.8; DB 11; Length 108;				
Best Local Similarity		87.6%; Pred. No. 0.0074;				
Matches		8; Conservative 0; Mismatches 12; Indels 0; Gaps 0;				
Qy	24035	GCCTGTAATCCAGCACTTTGGGAGCTGGAGTGGTGGATCAGAGGTCAGGAGATCAA	24094			
Db	1	GCCTGTAATCCAGCACTTTGGGAGGCGAAGCGGAGGATCACAAAGTCAGGAGATCGA	60			
Qy	24095	GACCATCGTGCCCAACATGGTGAACCCCGTCTCTAC	24131			
Db	61	GACCATCGTGCTAACATGGTGAACCCCGTCTTCC	97			
RESULT	7	HUMALCE221	103 bp	ss-RNA	PRI	15-APR-1994
LOCUS		Human carcinoma cell-derived Alu RNA transcript, clone CE221.				
DEFINITION						
ACCESSION		M87896				
VERSION		M87896.1	GI:174874			
KEYWORDS		Alu repeat.				
SOURCE		Homo sapiens male embryo carcinoma cDNA to other RNA.				
ORGANISM		Homo sapiens				
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE		1 (bases 1 to 103)				
AUTHORS		Sinnott,D., Richer,C., Deragon,J.-M. and Labuda,D.				
TITLE		Alu RNA transcripts in human embryonal carcinoma cells. Model of post-transcriptional selection of master sequences				
JOURNAL		J. Mol. Biol. (1992) In press				
FEATURES		Location/Qualifiers				
source		1..103				
		/organism="Homo sapiens"				
		/db_xref="taxon:9606"				
		/cell_line="Ntera2D1"				
		/dev_stage="embryo"				
		/sex="male"				

```

/tissue_type="carcinoma"
BASE COUNT      25 a      27 c      33 g      18 t
ORIGIN

Query Match      0.3%; Score 76; DB 9; Length 103;
Best Local Similarity 85.0%; Pred. No. 0.014;
Matches 85; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 24167 TCCTGTATCCAGCTACTCAAGAGGTGAGGCGAGGGAATGCTTGAACCGGGAGGT 24226
|||||
Db 4 TCCTGTATCCAGCTACTCAAGGGAAGTAAGGCGAGGAGAAATCGCTTGAACCGGGAGGC 63
|||||

QY 24227 GGACATTCGAGTGCAGTGCAGCACCTACACTACCTCCAG 24266
|||||
Db 64 GGAGGTTGCAGTGACCGGAGAGATCGTGCCATTCGCACTCCAG 103
|||||

RESULT 8
LOCUS HSU67808 108 bp RNA PRI 01-AUG-1997
DEFINITION Human small cytoplasmic Alu transcript.
ACCESSION U67808
VERSION U67808.1 GI:2289922
KEYWORDS Alu.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 108)
AUTHORS Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE cDNAs derived from primary and small cytoplasmic Alu (scAlu)
transcripts
JOURNAL J. Mol. Biol. 271 (2), 222-234 (1997)
MEDLINE 97415756
REFERENCE 2 (bases 1 to 108)
AUTHORS Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The
Children's Hospital of Philadelphia, 1004F Abramson Research
Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA
FEATURES
source
Location/Qualifiers
1..108
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TscAlu7"
1..108
/note="scAlu"
/rpt_family="Alu"
/rpt_type="dispersed"
BASE COUNT      22 a      37 c      28 g      21 t
ORIGIN

repeat_region 1..108
DNA
PRI 07-JAN-1995
HUMDLRAL/c HUMDLRAL 97 bp DNA
DEFINITION Human low density lipoprotein receptor gene, intron 4 (partial).
ACCESSION M14178
VERSION M14178.1 GI:187097
KEYWORDS low density lipoprotein receptor-1.
SEGMENT 1 of 2
SOURCE Human white blood cell DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 97)
AUTHORS Hobbs,H.H., Brown,M.S., Goldstein,J.L. and Russell,D.W.
TITLE Deletion of exon encoding cysteine-rich repeat of low density
lipoprotein receptor alters its binding specificity in a subject
with familial hypercholesterolemia
JOURNAL J. Biol. Chem. 261 (28), 13114-13120 (1986)
MEDLINE 87008518
COMMENT Analysis of the LDL-receptor gene of a patient with familial
hypercholesterolemia (FH) revealed the deletion of exon 5 resulting
from a homologous recombination between repetitive Alu sequences of
intron 4 and intron 5.
FEATURES
source
Location/Qualifiers
1..97
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="19p13.3"
<1..>97
/gene="LDLR"

introns

Query Match      0.3%; Score 75.4; DB 13; Length 103;
Best Local Similarity 87.2%; Pred. No. 0.017;
Matches 82; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 11305 CTGTAGATGGGTTTCGCCATGTTGCCAGGCTGTGCCAACTCTCGACCTGAAGTG 11364
|||||
Db 1 CAGTGTAGATAGGTTTCACCATGTTGCCAGGCTGTGCCAACTCTCGACCTGAAGTG 60
|||||

QY 11365 TTCACCCACCTCGCCCTCCCAAGTCTGGATT 11398
|||||
Db 61 ATCCACCCACCTCGACCTCCCAAGTCTGCGAAT 94
|||||

RESULT 10
HUMDLRAL/c HUMDLRAL 97 bp DNA
DEFINITION Human low density lipoprotein receptor gene, intron 4 (partial).
ACCESSION M14178
VERSION M14178.1 GI:187097
KEYWORDS low density lipoprotein receptor-1.
SEGMENT 1 of 2
SOURCE Human white blood cell DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 97)
AUTHORS Hobbs,H.H., Brown,M.S., Goldstein,J.L. and Russell,D.W.
TITLE Deletion of exon encoding cysteine-rich repeat of low density
lipoprotein receptor alters its binding specificity in a subject
with familial hypercholesterolemia
JOURNAL J. Biol. Chem. 261 (28), 13114-13120 (1986)
MEDLINE 87008518
COMMENT Analysis of the LDL-receptor gene of a patient with familial
hypercholesterolemia (FH) revealed the deletion of exon 5 resulting
from a homologous recombination between repetitive Alu sequences of
intron 4 and intron 5.
FEATURES
source
Location/Qualifiers
1..97
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="19p13.3"
<1..>97
/gene="LDLR"

introns

```



Query Match	0.3%;	Score 74.6;	DB 10;	Length 108;
Best Local Similarity	81.9%;	Pred. No. 0.023;		
Matches	86;	Conservative	0;	Mismatches 19; Indels 0; Gaps 0;
QY 24141	AAAAATTAGCAGGCATGGTAGACACATGCTGTAAATCCAGCTACTCAAGAGGCTGAGGC	24200		
Db	106 AAAAATTAGCAGGCGTGGTGGCAGGTGCTGTAAATCCAGCTACTCGGAGGCTGAGGC	47		
QY 24201	AGGGGAATTCCTTGAACCCGGGAGGTGGACATTGCAGTGAGCTGA	24245		
Db	46 AGGAAATGCTTTGAACCCAGGAGGACAGGTTGTGCTGAGGCCA	2		
RESULT 14				
HSLLDRD2				
LOCUS	HSLLDRD2	108 bp	DNA	PRI 20-MAY-1992
DEFINITION	Human LDL-receptor mutated gene with Intron 14 deletion junction.			
ACCESSION	X05251			
VERSION	X05251.1	GI:34336		
KEYWORDS	Alu repetitive sequence; low density lipoprotein receptor.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
AUTHORS	1 (bases 1 to 108) Horsthemke,B., Beisiegel,U., Dunning,A., Havinga,J.R., Williamson,R. and Humphries,S.			
TITLE	Unequal crossing-over between two alu-repetitive DNA sequences in the low-density-lipoprotein-receptor gene. A possible mechanism for the defect in a patient with familial hypercholesterolaemia			
JOURNAL	Eur. J. Biochem. 164 (1), 77-81 (1987)			
MEDLINE	87161901			
COMMENT	*source: hypercholesterol aemia See X05250 for corresponding normal gene sequence In the defective LDL-receptor gene the deletion occurred between two alu-repetitive sequences, that are in the same direction, the deletion eliminates exons 13 and 14 and changes the reading frame of the resulting spliced mRNA. Data kindly reviewed (07-DEC-1987) by HUMPHRIES S.			
FEATURES	Location/Qualifiers			
source	1..108 /organism="Homo sapiens" /db_xref="taxon:9606" /cell_type="blood leukocytes from a patient with familial"			
intron	1..108 /note="intron XIV fragment"			
BASE COUNT	28 a 20 c 40 g 20 t			
ORIGIN				
Query Match	0.3%;	Score 74.6;	DB 10;	Length 108;
Best Local Similarity	81.9%;	Pred. No. 0.023;		
Matches	86;	Conservative	0;	Mismatches 19; Indels 0; Gaps 0;
QY 24141	AAAAATTAGCAGGCATGGTAGACACATGCTGTAAATCCAGCTACTCAAGAGGCTGAGGC	24200		
Db	3 AAAAATTAGCAGGCGTGGTGGCAGGTGCTGTAAATCCAGCTACTCGGAGGCTGAGGC	62		
QY 24201	AGGGGAATTCCTTGAACCCGGGAGGTGGACATTGCAGTGAGCTGA	24245		
Db	63 AGGAAATGCTTTGAACCCAGGAGGACAGGTTGTGCTGAGGCCA	107		
RESULT 15				
HSU67807				
LOCUS	HSU67807	110 bp	RNA	PRI 01-AUG-1997
DEFINITION	Human small cytoplasmic Alu transcript.			
ACCESSION	U67807			
VERSION	U67807.1	GI:2289921		
KEYWORDS	Alu.			
SOURCE	human.			

```

ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 110)
AUTHORS      Shaikh,T.H., Roy,A.M., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE        cDNAs derived from primary and small cytoplasmic Alu (scAlu)
              transcripts
JOURNAL       J. Mol. Biol. 271 (2), 222-234 (1997)
MEDLINE       97415756
REFERENCE     2 (bases 1 to 110)
AUTHORS      Shaikh,T.H., Kim,J., Batzer,M.A. and Deininger,P.L.
TITLE        Direct Submission
JOURNAL       Submitted (22-AUG-1996) Human Genetics and Molecular Biology, The
              Children's Hospital of Philadelphia, 1004F Abramson Research
              Center, 34th and Civic Center Blvd., Philadelphia, PA 19104, USA
FEATURES
source       1..110
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="TscAlu6"
repeat_region 1..110
              /note="scAlu"
              /rpt_family="Alu"
BASE COUNT   26 a 39 c 24 g 21 t
ORIGIN
Query Match      0.3%; Score 74.4; DB 11; Length 110;
Best Local Similarity 84.0%; Pred. No. 0.025;
Matches 84; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 13530 GCCTGTAATCCACGACTTTGGGAGGCGCAAGCGGAGGATCACTTGAAGCCAGGAGTTC 13589
Db      1 GCCTGTAATCCACGACTTTGAGAGGCCAAGTGGGTGGATCACTTGAGCCAGGAGTTC 60
QY 13590 AAGACCGAGCTGGCCCAACATGGCAAAACCCCTGGCTCTACC 13629
Db      61 AAGACCGAGCTGTCAAACTGGTGAAGACCCCATCTTTCCC 100

```

Search completed: June 16, 2000, 20:09:08  
Job time: 199832 sec













PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PI (OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI: 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1: Page 1720; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 100 BP; 28 A; 22 C; 25 G; 22 T;

Query Match 0.2%; Score 62.4; DB 1; Length 100;  
Best Local Similarity 75.8%; Pred. NO. 1;  
Matches 75; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 16553 TTTTITTTTTCAGATGGAGTCTTACTGTGCTGCCTCAAGCTGAGTGGCGACAATC 16612  
DB 100 TTTGTTTGTTCACAGAGTGTCTACTGTGTCACCAGGCGNGAGTGAAGGTGCAATC 41

QY 16613 TCAGCTCACTGCAACCTCTGCTTCTTGGGTTTCAAGCAAT 16651  
DB 40 TCAGCTNATTGCAAAATTCCTGCTCCAGGTTCAAGCGAT 2

RESULT 8  
ID X12086/c  
AC X12086;  
DT 30-MAR-1999 (first entry)  
DE Human biallelic polymorphic DNA fragment EST98276b.  
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;  
KW detection; phenotypic typing; characteristic; infection; hereditary;  
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;  
KW treatment; marker; ss.  
OS Homo sapiens.  
PN W09820165-A2.  
PD 14-MAY-1998.  
PF 05-NOV-1997; U20313.  
PR 06-NOV-1996; US-030455.  
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
PI Hudson T, Lander ES, Wang D;  
DR WPI: 98-286974/25.  
PT New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease  
PS Claim 1; Page 219; 310pp; English.  
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
CC markers which have been isolated using the primers represented in  
CC X09121-X10268. The base occupying the polymorphic site is indicated by  
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as aquamaglobulinemia, diabetes insipidus, Lesch-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary

CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.  
SQ Sequence 100 BP; 21 A; 25 C; 22 G; 31 T;

Query Match 0.2%; Score 62.2; DB 1; Length 100;  
Best Local Similarity 76.8%; Pred. NO. 1.1;  
Matches 76; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 3638 GTGGCTCAAGCCTGTAAATCCCAACACTTTGGGAGGCTTAGTGGGAGGATTCCTTGAGCC 3697  
DB 99 GTGACTCACACCTATAATCTTGGCACTTTTAGGAGCTKAGGAAGGAGGATTTCTTGAAC 40

QY 3698 CAGTAGTTCAACAGCAGCTGGCCACATGGAGAACCC 3736  
DB 39 CAGGAGCTCAAGACCATCTCTGGGAAACATAGCAAGACTC 1

RESULT 9  
T25854  
ID T25854 standard; cDNA to mRNA; 91 BP.  
AC T25854;  
DT 22-OCT-1996 (first entry)  
DE Human gene signature HUMGS08084.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN W09514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATS/) MATSUBARA K.  
PI (OKUB/) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI: 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1; Page 1944; 2245pp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 91 BP; 18 A; 22 C; 28 G; 18 T;

Query Match 0.2%; Score 61; DB 1; Length 91;  
Best Local Similarity 78.7%; Pred. NO. 1.6;  
Matches 70; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 13704 ATCAGTTCAACCGGGAGCGAGAGGTTCCAGTGCAGCTGAGATTTCGCCACACACTACAG 13763  
DB 2 ATCAGTTGAGCCTAGGAGGAGGNGGTTCAAGTGCAGCTGAGATGGCACTCCTCGCTCCAG 61

```
QY 13764 CCTGGGTGACAGAGAGATTCGTCTCA 13792
    ||| ||| ||| ||| ||| ||| ||| |||
DB 62 CCTNGGTGACAGCGTGAGANNCTGTCTCA 90

RESULT 10
T20927/c
ID T20927 standard; cDNA to mRNA; 103 BP.
AC T20927.
DT 24-JUL-1996 (first entry)
DE Human gene signature HUMGS02180.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 758-759; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 103 BP; 22 A; 27 C; 21 G; 31 T;

Query Match 0.2%; Score 61.2; DB 1; Length 103;
Best Local Similarity 75.0%; Pred. No. 1.5;
Matches 75; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 13607 CATGGCAAAACCTGGCTCTACCAAAAATACACAAATTAGCTGGGATTTGGCACATGC 13666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 100 CATGGAGAATACTGTCCTCCATCTNAAAATACNAAAATCAGCTGGACATGTTGGCACAC 41

QY 13667 CTCTAATCCCAAGCTACTTGGAGGCTGAAGCAACAAGATC 13706
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 40 CTGTACCCACAGCTACTTGGAGGCTGAAGTGGAGGATC 1

RESULT 11
T26828
ID T26828 standard; cDNA to mRNA; 108 BP.
AC T26828.
DT 14-NOV-1996 (first entry)
DE Human gene signature HUMGS09078.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
```

```
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 2182; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 108 BP; 18 A; 33 C; 23 G; 28 T;

Query Match 0.2%; Score 61; DB 1; Length 108;
Best Local Similarity 73.8%; Pred. No. 1.6;
Matches 76; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 28509 GATCTCTTACACCTTGATCCACCCGCCCTCAGCTCCCAAGTGCAGGATTACAGGCAT 28568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 GATCTCTTACACCTTGATCCACCCGCCCTCAGCTCCCAAGTGCAGGATTACAGGCAT 60

QY 28569 GAGCCACCGTGCACGCCCTCTTTTCTTTTCTTTATTAAGACAAG 28611
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 GAGCCACCGTGCACGCCCTCTTTTCTTTTCTTTATTAAGACTGTACAGG 103

RESULT 12
T26213
ID T26213 standard; cDNA to mRNA; 103 BP.
AC T26213.
DT 13-NOV-1996 (first entry)
DE Human gene signature HUMGS08452.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 2029; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
```





PT New nucleic acid probes - have a labelled low frequency  
PT repetitive sequence for detecting overlaps among cloned DNA  
PS Disclosure; Page 8; 41pp; English.  
CC Genomic human placental DNA was mixed with a pair of PCR primers  
CC (see Q29014 and Q29015). The amplified DNA products were separated  
CC on an agarose gel and fragments in the 300-1000bp region were  
CC isolated. The fragments were ligated to M13mp19 RF DNA from which  
CC the 12bp SalI-BamHI insert had been removed. Competent JM109 were  
CC transformed by the ligation mixture and transformants were plated  
CC on NZY plates contg. beta-galactosidase indicator dye. Duplicate  
CC filter replicates were screened with two probes, one taken from the  
CC internal region of an Alu repeat (see Q29016) and the other a 5kb  
CC fragment contg. an L1 sequence from the region 5' to the human  
CC gamma-globin gene. Phage plaques which did not hybridise to either  
CC probe and did not react with the dye indicator were selected.  
CC Single-stranded DNA was extracted from them to isolate low-frequency  
CC repeat sequence probes LF12, LF15, LF16, LF17, LF18 and LF22.  
CC See Q29013-Q29017 and Q29021-Q29038.  
SQ Sequence 69 BP; 19 A; 18 C; 21 G; 11 T;

Query Match 0.2%; Score 58.4; DB 1; Length 69;  
Best Local Similarity 91.2%; Pred. No. 3.5;  
Matches 62; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 13544 CACTTTGGGAGGCCAAGCGCGGATCACTTGAAGCCAGGAGTTCAAGACACGCTGCC 13603  
|||||  
Db 1 CACTTTGGGAGGCCAAGCGCGGATCACTTGAAGCCAGGAGTTCAAGACACGCTGCC 60  
|||||

Qy 13604 CACATGG 13611  
|||||  
Db 61 CACATGG 68

Search completed: June 16, 2000, 21:16:14  
Job time: 202685 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2000, 03:46:29 ; Search time 8516.13 seconds  
(without alignments)  
13802.928 Million cell updates/sec

Title: US-08-852-495c-1\_copy\_140000\_169000  
Perfect score: 29001  
Sequence: 1 CCCGCCAATCCCATATGCAC.....TGTATCACAGAAAGTTACC 29001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues 156056  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 10  
Maximum DB seq length: 110

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :	EST:*
	1: em_est1:*
	2: em_est2:*
	3: em_est3:*
	4: em_est4:*
	5: em_est5:*
	6: em_est6:*
	7: em_est7:*
	8: em_est8:*
	9: em_est9:*
	10: em_est10:*
	11: em_est11:*
	12: em_est12:*
	13: em_est13:*
	14: em_est14:*
	15: em_est15:*
	16: em_est16:*
	17: em_est17:*
	18: em_est18:*
	19: em_est19:*
	20: gb_est1:*
	21: gb_est2:*
	22: gb_est3:*
	23: gb_est4:*
	24: gb_est5:*
	25: gb_est6:*
	26: gb_est7:*
	27: gb_est8:*
	28: gb_est9:*
	29: gb_est10:*
	30: gb_est11:*
	31: gb_est12:*
	32: gb_est13:*
	33: gb_est14:*
	34: gb_est15:*
	35: gb_est16:*
	36: gb_est17:*
	37: gb_est18:*
	38: gb_est19:*
	39: gb_est20:*
	40: gb_est21:*
	41: gb_est22:*
	42: gb_est23:*
	43: gb_est24:*
	44: gb_est25:*
	45: gb_est26:*
	46: gb_est27:*
	47: gb_est28:*
	48: gb_est29:*
	49: gb_est30:*
	50: gb_est31:*
	51: gb_est32:*
	52: em_est20:*
	53: em_est21:*
	54: em_est22:*
	55: em_est23:*
	56: em_est24:*
	57: em_est25:*
	58: em_est26:*
	59: gb_est33:*
	60: gb_est34:*
	61: gb_est35:*
	62: gb_est36:*
	63: gb_est37:*
	64: gb_est38:*
	65: em_est27:*
	66: em_est28:*
	67: em_est29:*
	68: em_est30:*
	69: gb_est39:*
	70: gb_est40:*
	71: gb_est41:*
	72: gb_est42:*
	73: gb_est43:*
	74: gb_est44:*
	75: em_est31:*
	76: em_est32:*
	77: em_est33:*
	78: em_est34:*
	79: gb_est45:*
	80: gb_est46:*
	81: gb_est47:*
	82: gb_gss1:*
	83: gb_gss2:*
	84: gb_gss3:*
	85: gb_gss4:*
	86: em_gss1:*
	87: em_gss2:*
	88: em_gss3:*
	89: em_gss4:*
	90: gb_gss5:*
	91: gb_gss6:*
	92: gb_gss7:*
	93: gb_gss8:*
	94: gb_gss9:*
	95: em_gss5:*
	96: em_gss6:*
	97: em_gss7:*
	98: em_gss8:*
	99: em_gss9:*
	100: em_gss10:*
	101: em_gss11:*
	102: gb_gss10:*
	103: gb_gss11:*
	104: em_gss12:*
	105: gb_gss12:*
	106: gb_gss13:*
	107: gb_gss14:*
	108: gb_gss15:*
	109: gb_gss16:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	%	Query
--------	---	-------

No.	Score	Match	Length	DB	ID	Description
1	93.8	0.3	109	30	AA243009	zr25h02.s
2	92.2	0.3	106	37	AA703692	ag81a10.r
3	91.6	0.3	105	105	AQ264176	CITBI-EI-
4	90.8	0.3	110	106	AQ386882	RPCI11-13
5	87.4	0.3	105	105	AQ282107	RPCI11-94
6	86	0.3	103	38	AA807640	nc08b05.s
7	85.4	0.3	103	108	AQ535244	RPCI-11-3
8	85	0.3	109	22	H11143	ym09c06.r1
9	85	0.3	110	39	AA897366	am06h02.s
10	85	0.3	110	94	AQ003188	RPCI11-ID
11	84.6	0.3	107	35	AA565533	nk42b11.s
12	84.4	0.3	109	30	AA244173	nc05h06.s
13	82.6	0.3	105	28	AA078003	7H12D08.C
14	82.8	0.3	110	30	AA244245	nc07a04.s
15	82.8	0.3	110	106	AQ386882	RPCI11-13
16	82.2	0.3	103	35	AA570438	nk63g02.s
17	82	0.3	103	84	B48914	RPCI11-4A12
18	82.2	0.3	103	108	AQ534922	RPCI-11-3
19	82	0.3	106	63	AT991750	wt48e01.x
20	81.8	0.3	109	94	AQ029690	RPCI11-41
21	82	0.3	110	32	AA369482	EST80906
22	81.2	0.3	104	108	AQ544583	CITBI-EI-
23	81.2	0.3	105	109	AQ637292	RPCI-11-4
24	81.2	0.3	106	30	AA250812	zs06a05.s
25	81.4	0.3	107	35	AA583252	nn41e04.s
26	81.4	0.3	108	84	B65160	CIT-HSP-201
27	80.6	0.3	103	108	AQ535244	nc06a05.s
28	80.6	0.3	106	38	AA812141	OB48h02.s
29	80.6	0.3	110	29	AA177157	nc02g07.s
30	80.2	0.3	101	35	AA583697	nn58f10.s
31	79.8	0.3	102	84	B48088	RPCI11-4N6
32	80	0.3	104	105	AQ321855	RPCI11-11
33	80.2	0.3	106	30	AA250812	zs06a05.s
34	80	0.3	107	33	AA385808	EST99495
35	80.2	0.3	108	32	AA370029	EST81584
36	80.2	0.3	109	84	B17434	345K2.TVB.C
37	80.2	0.3	109	84	B17434	345K2.TVB.C
38	80.4	0.3	110	30	AA244245	nc07a04.s
39	79.8	0.3	103	108	AQ584425	RPCI-11-4
40	79.6	0.3	106	20	T55212	YD43g11.s1
41	79.8	0.3	107	62	AI933497	wm74d02.x
42	79.8	0.3	109	94	AQ029690	RPCI11-41
43	79.2	0.3	100	30	AA252633	zq43g05.r
44	79.4	0.3	106	44	AT249096	qh73g09.x
45	79.4	0.3	106	108	AQ544957	CITBI-EI-

## ALIGNMENTS

```

RESULT 1
AA243009/c
LOCUS
DEFINITION
  zr25h02.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens
  cDNA clone IMAGE:664467 3' similar to contains Alu repetitive
  element;contains element LTR1 repetitive element,, mRNA sequence.
ACCESSION
  AA243009
VERSION
  AA243009.1 GI:1873869
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 (bases 1 to 109)
  Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
  Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
  Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
  Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  WashU-NCI human EST Project
  Unpublished (1997)
  On Dec 3, 1996 this sequence version replaced gi:1126869.

```

```

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1127 Std Error: 0.00
Seq primer: -41ml3 fwd. ST from Amersham
High quality sequence stop: 102.
Location/Qualifiers
  1. 109
  /organism="Homo sapiens"
  /db_xref="GDB:5426481"
  /db_xref="taxon:9606"
  /clone="IMAGE:664467"
  /clone_lib="Stratagene NT2 neuronal precursor 937230"
  /tissue_type="neuroepithelial cells"
  /dev_stage="Ntera-2 neuroepithelial cells"
  /lab_host="SOLR (kanamycin resistant)"
  /note="Organ: brain; Vector: pBluescript SK-; Site:1:
  EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:
  Oligo dT. Uninduced, exponentially growing neuroepithelial
  cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;
  Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG
  3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT 19 a 30 c 30 g 30 t
ORIGIN
Query Match 0.3%; Score 93.8; DB 30; Length 109;
Best Local Similarity 93.3%; Pred. No. 0.062; 7; Indels 0; Gaps 0;
Matches 98; Conservative 0; Mismatches 0;
QY 24032 CACGCTGTAAATCCAGCACCTTTGGGAGGCTGAGTGGGTGAATCAGGAGTCAGGAGAT 24091
|||||
Db 109 CACGCTGTAAATCCAGCACCTTTGGGAGGCTGAGTGGGTGAATCAGGAGTCAGGAGAT 50
QY 24092 CAAGACCATCTCGGCCACATGCTGAACCCCGTCTCTACTATAAA 24136
|||||
Db 49 CAAGACCATCTCGGTAAACACGGTGAACCCCGTCTCTACTATAAA 5
RESULT 2
AA703692/c
LOCUS
DEFINITION
  ag81a10.r1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone
  IMAGE:1140858 5' similar to contains Alu repetitive element,, mRNA
  sequence.
ACCESSION
  AA703692
VERSION
  AA703692.1 GI:2713610
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 (bases 1 to 106)
  Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
  Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
  Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
  Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  WashU-NCI human EST Project
  Unpublished (1997)
  On Sep 12, 1996 this sequence version replaced gi:1397630.
  Contact: Wilson RK
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: est@watson.wustl.edu
  This clone is available royalty-free through LNL ; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.

```

Seq primer: -28ml3 rev1 ET from Amersham  
High quality sequence stop: 53.

# FEATURES

Location/Qualifiers  
1..106  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1140858"  
/clone\_lib="Stratagene hMT neuron (#937233)"  
/dev\_stage="hMT neurons"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Differentiated, post mitotic hMT neurons. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."  
19 a 29 c 29 g 29 t

# BASE COUNT

ORIGIN

Query Match 0.3%; Score 92.2; DB 37; Length 106;  
Best Local Similarity 92.4%; Pred. No. 0.096; Mismatches 0; Indels 0; Gaps 0;  
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24032 CACGCTGTATCCAGCAGCTTTGGAGGCTGAGGTGGTGAATCAGCAGGTCAGGAGAT 24091  
|||||  
DB 106 CAGCGTGTATCCAGCAGCTTTGGAGGCTGAGGCGGCAGATCAGGTCAGGAGAT 47  
|||||

QY 24092 CAAGACCATCTCGGCCAACATGTTGAACCCCGTCTCTACTAAA 24136  
|||||  
DB 46 CGAGACCATCTCGGTGAACAGCGTGAACCTCGTCTCTACTAAA 2

# RESULT 3

AQ264176 106 bp DNA GSS 27-OCT-1998  
LOCUS CITBI-E1-2509A2.TF CITBI-E1 Homo sapiens genomic clone 2509A2,  
DEFINITION genomic survey sequence.  
ACCESSION AQ264176  
VERSION AQ264176.1 GI:3792743  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 106)  
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.  
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready Map Building  
JOURNAL Unpublished (1998)  
COMMENT Other\_GSSs: CITBI-E1-2509A2.TR  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/tldb/hungen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: M13-21  
Class: BAC ends.  
Location/Qualifiers  
1..106  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2509A2"  
/clone\_lib="CITBI-E1"  
/sex="male"  
/cell\_type="sperm"

# FEATURES

Location/Qualifiers  
1..106  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2509A2"  
/clone\_lib="CITBI-E1"  
/sex="male"  
/cell\_type="sperm"

Query Match 0.3%; Score 90.8; DB 106; Length 110;  
Best Local Similarity 89.1%; Pred. No. 0.14; Mismatches 0; Indels 0; Gaps 0;  
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9524 GGTTTTCACCATGTTGCCAGGCTGGTGTGAACCTCCTGACCTCAAGTGATCCACCTGCC 9583

/note="Vector: pBelOBAC11; Site\_1: EcoRI; Site\_2: EcoRI;  
CalTech Human BAC Library D"  
BASE COUNT 25 a 30 c 34 g 17 t  
ORIGIN

Query Match 0.3%; Score 91.6; DB 105; Length 106;  
Best Local Similarity 91.5%; Pred. No. 0.11; Mismatches 0; Indels 0; Gaps 0;  
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24017 CCGGTGTCAGTGGCTCACCCCTGTATCCAGCAGCTTTGGAGGTCAGGTGGTGAATC 24076  
|||||  
DB 1 CCGGCGCCAGAGTCTCACCCCTGTATCCAGCAGCTTTGGAGCAGCGCGGTGGATC 60  
|||||

QY 24077 ACGAGTCAGGAGATCAAGACCATCTCGGCCAACATGTTGAACCC 24122  
|||||  
DB 61 ACGAGTCAGGAGATCAAGACCCCTCTGCTGTACATGGTGAACCC 106

# RESULT 4

AQ386882/c 110 bp DNA GSS 21-MAY-1999  
LOCUS RPC11-134I4.TV RPCI-11 Homo sapiens genomic clone RPCI-11-134I4,  
DEFINITION genomic survey sequence.  
ACCESSION AQ386882  
VERSION AQ386882.1 GI:4357905  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 110)  
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.  
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building  
JOURNAL Unpublished (1997)  
COMMENT Other\_GSSs: RPC11-134I4.TJ  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@tigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@tigr.org), please contact Pieter de Jong (pieter@tigr.org) or BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/tldb/hungen/bac\_end\_search/bac\_end\_search.html  
Seq primer: T7  
Class: BAC ends.  
Location/Qualifiers  
1..110  
/organism="Homo sapiens"  
/db\_xref="GDB:7551267"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-134I4"  
/clone\_lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACe3 6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPC11 Human Male BAC Library"

# FEATURES

Location/Qualifiers  
1..110  
/organism="Homo sapiens"  
/db\_xref="GDB:7551267"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-134I4"  
/clone\_lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACe3 6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPC11 Human Male BAC Library"

# BASE COUNT

26 a 26 c 38 g 20 t  
ORIGIN

Query Match 0.3%; Score 90.8; DB 106; Length 110;  
Best Local Similarity 89.1%; Pred. No. 0.14; Mismatches 0; Indels 0; Gaps 0;  
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
|||||
Db 110 GGGTTTACCATTGTTCCAGGCTGGTCTTGAACCTTGACCTCAAGCATCCACCTGCC 51
QY 9584 TCAGCCTCACATGTTCTGGGATTACAGCGGTGAGCCACCATGCTGGCC 9633
|||||
Db 50 TCAGCCTCCCAAAGTACTTGGATTACAGCGGTGAGCCACTGCTCCCGGCC 1
|||||

RESULT 5
AQ282107 AQ282107 105 bp DNA GSS 27-APR-1999
LOCUS RPCI11-94B21.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-94B21,
DEFINITION genomic survey sequence.
ACCESSION AQ282107
VERSION AQ282107
KEYWORDS GSS.
SOURCE AQ282107.1 GI:3907976
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 105)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
TITLE Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
JOURNAL Use of human BAC End Sequences for Sequence-Ready Map Building
COMMENT Unpublished (1998)
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..105
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RPCI-11-94B21"
/sex="Male"
/cell_type="Lymphocytes"
/notes="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"
BASE COUNT 26 a 31 c 30 g 18 t
ORIGIN
Query Match 0.3%; Score 87.4; DB 105; Length 105;
Best Local Similarity 89.5%; Pred. No. 0.34;
Matches 94; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 24029 GCTCAGCGCTGTAATCCAGCAGCTTTGGAGGCTGAGGTGGGTGAATCAGCAGGTCAGGA 24088
|||||
Db 1 GCTCAGCGCTGTAATCCAGCAGCTCTGGAGGCCAAGGTGGGTGATCAGCAGGGCATGA 60
|||||

QY 24089 GATCAGACCATCCYGGCCACATGGTGAACCCCGTCTCTACTA 24133
|||||
Db 61 GTACGAGACCGCTCGACCAACATGGTGAACCCCGTCTCTACTA 105
|||||

RESULT 6
AA807640/c AA807640 103 bp mRNA EST 05-MAR-1998
LOCUS nx08b05.s1 NCI_CGAP_GC3 Homo sapiens cDNA clone IMAGE:1255473 3'
DEFINITION similar to contains Alu repetitive element;; mRNA sequence.
```

```
AA807640
VERSION AA807640.1 GI:2877108
EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 103)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
On Jan 19, 1998 this sequence version replaced gi:2151346.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 774 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 87.
Location/Qualifiers
1..103
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1255473"
/clone_lib="NCI_CGAP_GC3"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/notes="Vector: p77T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled
germ cell tumors, and was then primed with a Not I -
oligo(dr) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified p77T3
vector. Library is not normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 19 a 27 c 30 g 27 t
ORIGIN
Query Match 0.3%; Score 86; DB 38; Length 103;
Best Local Similarity 90.2%; Pred. No. 0.5;
Matches 92; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 24031 TCACGCCCTGTAATCCAGCAGCTTTGGGAGGCTGAGGTGGGTGAATCAGCAGGTCAGGAGA 24090
|||||
Db 103 TCACACCTGTAATCCAGCAGCTTTGGGAGGCGCGAGGTGAGCGGATCACAAAGTCAGGAGA 44
|||||

QY 24091 TCAAGACCATCTCGCCACATGGTGAACCCCGTCTCTACT 24132
|||||
Db 43 TCGAGACCATCTGCTGTACACGGGTGAACCCCGTCTCTACT 2
|||||

RESULT 7
AQ535244 AQ535244 103 bp DNA GSS 18-MAY-1999
LOCUS RPCI-11-317H22.TV RPCI-11 Homo sapiens genomic clone
DEFINITION RPCI-11-317H22, genomic survey sequence.
ACCESSION AQ535244
VERSION AQ535244.1 GI:4846934
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
```







```
BASE COUNT      22 a      34 c      26 g      25 t
ORIGIN

Query Match      0.3%; Score 84.6; DB 35; Length 107;
Best Local Similarity 86.9%; Pred. No. 0.72;
Matches 93; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 13657 TGGCAGCATCCCTGTATCCAGCTACTCTGGGAGGCTGAAGCACAGAATCACTTGAACCG 13716
|||||
Db 107 TGGTGTGTCCTGTATCCAGCTACTCTAGGAGGCTGAGGCACGAGAAATCACTTGAACCT 48
|||||

QY 13717 GGGAGGACAGAGGTTGCAGTGAGCTGAGATTTGCCCACTACACTACAG 13763
|||||
Db 47 GGGAGGACAGGCTTGCAGTGAGCTGAGATTTGAGCCACTGCACCTCCAG 1

RESULT 12
AA244173/c
LOCUS      AA244173      109 bp      mRNA      EST      20-AUG-1997
DEFINITION nc05h06.s1 NCI-CGAP_Prl Homo sapiens cDNA clone IMAGE:1007291
similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION  AA244173
VERSION     AA244173.1 GI:1874876
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 109)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   On Nov 29, 1993 this sequence version replaced gi:430513.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
M.D., Michael Emerit-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
DNA Sequencing by: Genome Systems Inc., Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -41m13 fwd. ET from Amersham
High quality sequence stop: 90.
Location/Qualifiers
1..109
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1007291"
/clone_lib="NCI-CGAP_Prl"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/Note="Vector: PAMPl0; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected, histologically normal
prostate epithelial cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into PAMPl0 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Krizman."
28 a      28 c      31 g      22 t

BASE COUNT
ORIGIN

Query Match      0.3%; Score 82.6; DB 28; Length 105;
Best Local Similarity 86.7%; Pred. No. 1.2;
Matches 91; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 13635 TACAACAATAGCTGGGCATTTGGCACATGCTGTAAATCCAGCTACTTTGGAGGCTGA 13694
|||||
Db 105 TACAAAATATAGCTGGGCATGTAGCGCAGCATGTAAATCCAGCTACTTTGGAGGCTGA 46
|||||

QY 13695 AGCACAAGATATCACTTGAACCCGGGAGGAGGAGTTGCAGTGAGC 13739
|||||
Db 45 GACCCCGAGATTCGTTGAACCCAGGAGGAGAGGTTGCAGTGAGC 1

Query Match      0.3%; Score 84.4; DB 30; Length 109;
Best Local Similarity 89.2%; Pred. No. 0.75;
Matches 91; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 9478 CCCTACACCATATGCCCGCTAATTTTGTATTTTAGTAGAGACAGGGTTTCACCATGT 9537
|||||
Db 109 CACCCACACCATATGCTGGCTAATTTCTGTATTTTAGTAGAGCAGGGTTTCACCATGT 50
|||||

QY 9538 TGGCCAGGCTGGTGTGTAACCTCTGACCTCAAGTGATCCACC 9579
|||||
Db 49 TGCCAGGCTGGTGCATGAACCTCTGACCTAGGTGATCCACC 8
|||||

RESULT 13
AA078003/c
LOCUS      AA078003      105 bp      mRNA      EST      24-SEP-1999
DEFINITION 7H12D08 Chromosome 7 HeLa cDNA Library Homo sapiens cDNA clone
7H12D08, mRNA sequence.
ACCESSION  AA078003
VERSION     AA078003.1 GI:1837477
KEYWORDS   EST.
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 105)
AUTHORS   Touchman, J.W., Bouffard, G.G., Weintraub, L.A., Idol, J.R., Wang, L.,
Robbins, C.M., Nusbaum, J.C., Lovett, M. and Green, E.D.
2006 expressed-sequence tags derived from human chromosome
7-enriched cDNA libraries
Genome Res. 7 (3), 281-292 (1997)
JOURNAL   97228905
MEDLINE
COMMENT   On Apr 14, 1993 this sequence version replaced gi:693433.
Contact: Eric D. Green
Genome Technology Branch
National Human Genome Research Institute/NIH
49 Convent Dr., MSC4431, Building 49, Room 2A08, Bethesda, MD 20892
Tel: 3014020201
Fax: 3014024735
Email: egreen@nhgri.nih.gov
Plate: 12 row: D column: 08
Seq primer: -21M13 (ABI).
Location/Qualifiers
1..105
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="7H12D08"
/sex="female"
/cell_line="HeLa cell line; ATCC"
/lab_host="E. coli strain DH5 alpha"
/Note="Vector: PAMPl0; cDNA was generated from cytoplasmic
RNA using a mixture of random DNA hexamers and oligo(dT).
From this pool of cDNA, human chromosome 7-enriched cDNA
was isolated by direct cDNA selection using chromosome 7
genomic DNA (cosmids). The resulting direct-selected cDNA
was cloned into a plasmid vector using a non-directional
uracil DNA glycosylase (UDG)-mediated cloning strategy."
20 a      33 c      23 g      29 t

BASE COUNT
ORIGIN

Query Match      0.3%; Score 82.6; DB 28; Length 105;
Best Local Similarity 86.7%; Pred. No. 1.2;
Matches 91; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 13635 TACAACAATAGCTGGGCATTTGGCACATGCTGTAAATCCAGCTACTTTGGAGGCTGA 13694
|||||
Db 105 TACAAAATATAGCTGGGCATGTAGCGCAGCATGTAAATCCAGCTACTTTGGAGGCTGA 46
|||||

QY 13695 AGCACAAGATATCACTTGAACCCGGGAGGAGGAGTTGCAGTGAGC 13739
|||||
Db 45 GACCCCGAGATTCGTTGAACCCAGGAGGAGAGGTTGCAGTGAGC 1
```

RESULT	14
AA244245	
LOCUS	110 bp mRNA EST 20-AUG-1997
DEFINITION	nc07a04.s1 NCI-CGAP_Prl Homo sapiens cDNA clone IMAGE:1007406 similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION	AA244245
VERSION	AA244245.1 GI:1875104
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 110)
JOURNAL	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)
COMMENT	On Jan 24, 1995 this sequence version replaced gi:634306. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 E-mail: Robert_Strausberg@nih.gov Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D., Michael Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: David B. Krizman, Ph.D. cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://www-bio.llnl.gov/bbrp/image/image.html">www-bio.llnl.gov/bbrp/image/image.html</a>

```

Seq primer: -41ml3 fwd.  ET from Amersham
High quality sequence stop: 90.
Location/Qualifiers
1..110
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1007406"
/clone_lib="NCI_CGAP_Pr1"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected, histologically normal
prostate epithelial cells. Double-stranded cDNA was
ligated to EcoRI adaptors. 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into pAMP10 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Krizman."
17 a 26 c 28 g 38 t 1 others

```

	Query Match	0.3%	Score 82.8	DB 30	Length 110
	Best Local Similarity	87.4%	Pred. No. 1.1		
	Matches 90	Conservative 0	Mismatches 13	Indels 0	Gaps 0
QY	16554	TTTTTTTTTGGAGATGGAGTCTTACTCTGTGCGCTCAAGCTGGAGTCGAGTGGGCACAATCT	16613		
Db	1	TTTTTTTTTGGAGATGGAGTCTTGATCTGTGCGCAGCTGGAGTCGAGTGGGCAGANTCT	60		
QY	16614	CAGCTCAGTCGAACCTCTGCCTTCTGGGTTCAAGCAATTTGCG	16656		
Db	61	TGCTCAGTCGAACCTCTGCCTCTGGGTTCAAGAGATTTTC	103		

RESULT 15  
AQ386882

LOCUS	AQ386882	110 bp	DNA	GSS	21-MAY-1999							
DEFINITION	RPC111-134I4.TV RPC1-11 Homo sapiens genomic clone RPC1-11-134I4, genomic survey sequence.											
ACCESSION	AQ386882											
VERSION	AQ386882.1	GI:4357905										
KEYWORDS	GSS.											
SOURCE	human.											
ORGANISM	Homo sapiens											
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;											
	Eutheria; Primates; Catarrhini; Hominidae; Homo.											
REFERENCE	1 (bases 1 to 110)											
AUTHORS	Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.											
TITLE	Use of BAC End Sequences from Library RPC1-11 for Sequence-Ready Map Building											
JOURNAL	Unpublished (1997)											
COMMENT	Other GSSs: RPC111-134I4.TJ											

JOURNAL  
COMMENT

Unpublished (1997)  
Other\_GSSs: RPC111-134I4.TJ  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbsctigr.org

Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics ([inforesgen.com](http://inforesgen.com)). BAC end search page: [http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: T7  
Class: BAC ends.

FEATURES	Location/Qualifiers
source	1..1110
	/organism="Homo sapiens"
	/db_xref="GBA:7551267"
	/db_xref="taxon:9606"
	/clone="RPCI-11-13414"
	/clone_lib="RPCI-11"
	/sex="Male"
	/cell_type="Lymphocytes"
	/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; RPC111 Human Male BAC Library"
BASE COUNT	26 a 26 c 38 g 20 t
ORIGIN	

	Query Match	0.3%	Score 82.8;	DB 106;	Length 110;
	Best Local Similarity	84.5%;	Pred. No. 1.1;		
	Matches 93;	Conservative	0;	Mismatches 17;	Indels 0;
	Gaps				
QY	13510	GGCCAGGTGTGGGTTCATGCTCTTAATCCCAAGCACTTTGGAGGCCAAGGCAGCGCGGA	13569		
Db	1	GGCCGGGAGCAGTGGCTCAGCCCTGTAAATCAAAGTACTTTGGAGGCTGAGGCAGGTGGA	60		
QY	13570	TCAGTTCAAGCCAGGAGGTTTCAAGACCACCGCTGCCCAACATGGCAAAACCC	13619		
Db	61	TCGCTTCAGGTTCAGGAGTTTCAAGACCAGCCTGGCAACAATGGTGAACACC	110		

Search completed: June 16, 2000, 15:09:23  
Job time: 181930 sec



